GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:20:59 ; Search time 10173.5 Seconds
(without alignments)
13311.631 Million cell updates/sec

Title: US-09-171-553B-3
Perfect score: 8209
Sequence: 1 gtggtgtacgactgtgggcc.....aaaaaaaaaaaaaaaaaaa 8209

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.	Score	Query Match	Length	DB	ID	Description
1	8209	100.0	8209	6	A66553	A66553 Sequence 3
2	7751	94.4	8196	6	A66552	A66552 Sequence 2
3	4833	58.9	8918	14	PEN133816	AJ133816 Porcine e
4	4640	56.5	8849	14	PEN133817	AJ133817 Porcine e
5	4540	55.3	8763	14	PEN293657	AJ293657 Porcine e
6	4532	55.2	8750	14	PEN133818	AJ133818 Porcine e
7	4096	49.9	7808	14	PERY17013	Y17013 Porcine end
8	2967	36.1	7333	4	AF038601	AF038601 Sus scrof
9	2967	36.1	7333	6	AR130474	AR130474 Sequence
10	2682	32.7	3320	6	A66551	A66551 Sequence 1
11	2034	24.8	3025	4	AF147808	AF147808 Sus scrof
12	1971	24.0	1971	14	AF014162	AF014162 Porcine e
13	1499	18.3	1974	6	A66559	A66559 Sequence 9
14	1466	17.9	3482	6	AX002804	AX002804 Sequence
15	1466	17.9	3482	14	PERENV2	Y12239 Porcine end
16	1064	13.0	8918	14	PEN293656	AJ293656 Porcine e
17	876	10.7	927	4	SSU77599	U77599 Sus scrofa
18	848	10.3	4402	6	AX052635	AX052635 Sequence
19	797	9.7	6076	6	AX052636	AX052636 Sequence
20	797	9.7	7362	6	AX052634	AX052634 Sequence
21	746	9.1	6076	6	AX052647	AX052647 Sequence
22	704	8.6	8132	6	AR130475	AR130475 Sequence
23	622	7.6	942	14	MLVRT10	X99933 Murine leuk
24	570	6.9	4918	6	AX052637	AX052637 Sequence
25	557	6.8	8132	4	AF038600	AF038600 Sus scrof
26	547	6.7	7873	6	AX052638	AX052638 Sequence
27	457	5.6	2462	6	AX002802	AX002802 Sequence
28	457	5.6	2462	14	PERENV1	Y12238 Porcine end
29	446	5.4	1042	14	PERY17012	Y17012 Porcine end
30	437	5.3	2000	14	PEN288588	AJ288588 Porcine e
31	375	4.6	8060	6	AR130473	AR130473 Sequence
32	375	4.6	8132	4	AF038599	AF038599 Sus scrof
33	367	4.5	1968	14	PEN288589	AJ288589 Porcine e
34	335	4.1	2000	14	PEN288592	AJ288592 Porcine e
35	321	3.9	746	14	PEN298073	AJ298073 Porcine e
36	318	3.9	590	14	PEN298074	AJ298074 Porcine e
37	276	3.4	276	4	SSAJ5400	AJ005400 Sus scrof
38	225	2.7	276	4	SSY18745	Y18745 Sus scrofa
39	223	2.7	276	4	SSAJ5409	AJ005409 Sus scrof
40	217	2.6	276	4	SSY18748	Y18748 Sus scrofa
41	201	2.4	276	4	SSY18749	Y18749 Sus scrofa
42	189	2.3	276	4	SSAJ5410	AJ005410 Sus scrof
43	161	2.0	810	14	AF033259	AF033259 Porcine e
44	158	1.9	522	14	AF033260	AF033260 Porcine e
45	158	1.9	927	4	SSU77600	U77600 Sus scrofa

ALIGNMENTS

RESULT 1
A66553
LOCUS A66553 8209 bp DNA PAT 29-MAR-1999
DEFINITION Sequence 3 from Patent WO9740167.
ACCESSION A66553
VERSION A66553.1 GI:4538106
KEYWORDS .
SOURCE porcine endogenous retrovirus.
ORGANISM porcine endogenous retrovirus
Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses; 1-Mammalian type C virus group.
REFERENCE 1 (bases 1 to 8209)
AUTHORS Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
TITLE PORCINE RETROVIRUS
JOURNAL Patent: WO 9740167-A 3 30-OCT-1997;
Q ONE BIOTECH LTD (GB)
FEATURES Location/Qualifiers
source 1. .8209
/organism="porcine endogenous retrovirus"
/db_xref="taxon:61673"

```
     CDS             588. .2162
                     /codon_start=1
                     /product="VIRAL CORE PROTEIN"
                     /protein_id="CAB39347.1"
                     /db_xref="GI:4538107"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPGSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSSSYLPRDRGAADLAGTPTCSPTPLSSTGCCEG
                     TSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREREEERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKD"
     CDS             <2163. .5747
                     /codon_start=1
                     /product="REVERSE TRANSCRIPTASE"
                     /protein_id="CAB39348.1"
                     /db_xref="GI:4538108"
                     /translation="GRRGSDPLPEPRVTLKVEGQPVEFLVDTGAEHSVLLQPLGKLKE
                     KKSWVMGATGQRQYPWTTRRTVDLGVGRVTHSFLVIPECPVPLLGRDLLTKMGAQISF
                     EQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLA
                     KQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLP
                     VRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCL
                     RLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQ
                     VTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLR
                     GGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKG
                     GFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRR
                     PVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPP
                     DRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVR
                     KDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELM
                     ALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLE
                     ALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPIIETPKAPEPRRQ
                     YTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEGLEYVQQIHRLTHLGTKHLQQL
                     VRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRIPPGKRLRGSHPGAHWEVDFTEVK
                     PAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKILEEIFPRFGIPKVIGSDNGPA
                     FVAQVSQGLAKILGIDWKLHCAYRPQSSGQVERMNRTIKETLTKLTTETGINDWMALL
                     PFVLFRVRNTPGQFGLTPYELLYGGPPPLAEIAFAHSADVLLSQPLFSRLKALEWVRQ
                     RAWKQLREAYSGGDLQVPHRFQVGDSVYVRRHRAGNLETRWKGPYLVLLTTPTAVKVE
                     GIPTWIHASHVKPAPPPDSGWRAEKTENPLKLRLHRLVPYSNNNSPGQ"
     CDS             5620. .7590
                     /codon_start=1
                     /product="VIRAL ENVELOPE PROTEIN"
                     /protein_id="CAB39349.1"
                     /db_xref="GI:4538109"
                     /translation="MHPTLSRRHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSK
                     RLIDSSNPHRPLSLTWLIIDPDTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTP
                     PNLVRSYGFYCCPGTEKEKYCGGSGESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVN
                     SGPGKYKMMKLYKDKSCSPSDLDYLKISFTERKTGKYSKVDKWYELGNSFLLYGGGAG
                     STLTIRLRIETGTEPPVAMGPDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSNS
                     TTGLIPTNTPRNSPGVPVKTGQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYYE
                     GMAKERKFNVTKEHRNQCTWGSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQA
                     SENQYLVPGYNRWWACNTGLTPCVSTSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEYD
                     YRYNRPKREPVSLTLAVMLGLGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLRA
                     LKESVSNLEESLTSLSEVVLQNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDSM
                     NKLRKKLERRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLINR
                     FVAFVRERVSAVQIMVLRQQYQGLLSQGETDL"
BASE COUNT     2168 a   2064 c   2154 g   1823 t
ORIGIN
```

```
  Query Match           100.0%;  Score 8209;  DB 6;  Length 8209;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 8209;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 60

Qy      61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAGACCGCTTCTCGTGAGTGATTTGGGGTGTCGCCTCTTCCGAGCCCGGACGAGGGGGAT 120

Qy     121 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctta 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGTTCTTTTACTGGCCTTTCATTTGGTGCGTTGGCCGGGAAATCCTGCGACCACCCCTTA 180

Qy     181 cacccgagaaccgacttggaggtaaagggatccccttttggaacgtgtgtgtgtgtcggcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CACCCGAGAACCGACTTGGAGGTAAAGGGATCCCCTTTGGAACGTGTGTGTGTGTCGGCC 240

Qy     241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGCGTCTCTGTTCTGAGTGTCTGTTTTCGGTGATGCGCGCTTTCGGTTTGCAGCTGTCCT 300

Qy     301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTCAGACCGTAAGGACTGGAGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAGGCTG 360

Qy     361 ccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCACCCTGGGGGACGCCCCGGGAGGTGGGGAGAGCCAGGGACGCCTGGTGGTCTCCTACT 420

Qy     421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCG 480

Qy     481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ACTCTTTTGCCTGCTTGTGGAAGACGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGTTT 540

Qy     541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CTGTCTCGTGTGTCTTTGTCTTGTGCGTCCTTGTCTACAGTTTTAATATGGGACAGACAG 600

Qy     601 tgactacccccctagtttgactctcgaccattggactgaagttagatccagggctcata 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TGACTACCCCCCTTAGTTTGACTCTCGACCATTGGACTGAAGTTAGATCCAGGGCTCATA 660

Qy     661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGCCTCTGAATGGCCAA 720

Qy     721 cattcgatgttggatggccatcagagggaccttaattctgaaattatcctggctgtta 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCTGAAATTATCCTGGCTGTTA 780

Qy     781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AGGCAATCATTTTTCAGACTGGACCCGGCTCTCATCCTGATCAGGAGCCCTATATCCTTA 840

Qy     841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CGTGGCAAGATTTGGCAGAAGATCCTCCGCCATGGGTTAAACCATGGCTAAATAAACCAA 900

Qy     901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GAAAGCCAGGTCCCCGAATCCTGGCTCTTGGAGAGAAAAACAAACACTCGGCCGAAAAAG 960

Qy     961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TCGAGCCCTCTTCCTCGTATCTACCCCGAGATCGAGGAGCCGCCGACTTGGCCGGAACCC 1020

Qy    1021 caacctgttcccccaccccctttatccagcacagggtgctgtgaggggacctctgcccctc 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CAACCTGTTCCCCCACCCCCTTATCCAGCACAGGGTGCTGTGAGGGGACCTCTGCCCCTC 1080

Qy    1081 ctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccaccc 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CTGGAGCTCCGGTGGTGGAGGGACCTGCTGCCGGGACTCGGAGCCGGAGAGGCGCCACCC 1140

Qy    1141 cggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgccag 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CGGAGCGGACAGACGAGATCGCGATATTACCGCTGCGCACCTATGGCCCTCCCATGCCAG 1200
```

```
Qy   1201 ggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgga 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGGA 1260

Qy   1261 aaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtccc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGAGTCCC 1320

Qy   1321 ttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttca 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTCA 1380

Qy   1381 caaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgacg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAGAAAAAATGTTCCTGGGGCCGACG 1440

Qy   1441 ggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggtt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGATTTCCCTTGACTCGCCCCGGTT 1500

Qy   1501 gggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtgg 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCTGGTGG 1560

Qy   1561 cgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAGGTGATGC 1620

Qy   1621 agggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcggt 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AGGGACCGAACGAACCTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGGT 1680

Qy   1681 tcaccccttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcattg 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TCACCCCTTTTGATCCTACCTCAGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTCATTG 1740

Qy   1741 ggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggctg 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GGCAGTCGGCTCTGGATATCAGGAAGAAACTTCAGAGACTGGAAGGGTTACAGGAGGCTG 1800

Qy   1801 agttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagagg 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AGTTACGTGATCTAGTGAGAGAGGCAGAGAAGGTGTATTACAGAAGGGAGACAGAAGAGG 1860

Qy   1861 agaaggaacagagaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacggc 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGAAGGAACAGAGAAAAGAAAAGGAGAGAAGAAAGGGAGGAAAGACGTGATAGACGGC 1920

Qy   1921 aagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggaga 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AAGAGAAGAATTTGACTAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAGA 1980

Qy   1981 gagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaata 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAATA 2040

Qy   2041 ggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagga 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 GGACCCCACTCGACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGGACACTGGGCAAGGA 2100

Qy   2101 actgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagatt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACTGCCCCAAGAAGGGAAACAAAGGACCGAAGGTCCTAGCTCTAGAAGAAGATAAAGATT 2160

Qy   2161 aggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtggaggggc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGAGGGGC 2220

Qy   2221 aaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccattag 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 AACCAGTTGAGTTCCTGGTTGATACCGGAGCGGAGCATTCAGTGCTGCTACAACCATTAG 2280

Qy   2281 gaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatccat 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GAAAACTAAAAGAAAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACGGCAGTATCCAT 2340

Qy   2341 ggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtca 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GGACTACCCGAAGAACCGTTGACTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTCA 2400

Qy   2401 tccctgagtgcccagtacccttctaggtagagacttactgaccaagatgggagctcaaa 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TCCCTGAGTGCCCAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAAA 2460

Qy   2461 tttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttga 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TTTCTTTTGAACAAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCACTGTGTTGA 2520

Qy   2521 ccctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagata 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CCCTCCAATTAGATGATGAATATCGACTATATTCTCCCCAAGTAAAGCCTGATCAAGATA 2580

Qy   2581 tacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttgg 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TACAGTCCTGGTTGGAGCAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGGTTTGG 2640

Qy   2641 caaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtca 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CAAAGCAAGTTCCCCCACAGGTTATTCAACTGAAGGCCAGTGCTACACCAGTATCAGTCA 2700

Qy   2701 gacagtacccttgagtagagaggctcgagaaggaatttggccgcatgttcaaagattaa 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 GACAGTACCCCTTGAGTAGAGAGGCTCGAGAAGGAATTTGGCCGCATGTTCAAAGATTAA 2760

Qy   2761 tccaacagggcatcctagttcctgtccaatccccttggaatactcccctgctaccggtta 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TCCAACAGGGCATCCTAGTTCCTGTCCAATCCCCTTGGAATACTCCCCTGCTACCGGTTA 2820

Qy   2821 ggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaaggg 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GGAAGCCTGGGACCAATGATTATCGACCAGTACAGGACTTGAGAGAGGTCAATAAAAGGG 2880

Qy   2881 tgcaggacatacacccaacggtcccgaacccttataacctcttgagcgccctcccgcctg 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TGCAGGACATACACCCAACGGTCCCGAACCCTTATAACCTCTTGAGCGCCCTCCCGCCTG 2940

Qy   2941 aacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacacc 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 AACGGAACTGGTACACAGTATTGGACTTAAAAGATGCCTTCTTCTGCCTGAGATTACACC 3000

Qy   3001 ccactagccaaccacttttttgccttcgaatggagagatccaggtacgggaagaaccgggc 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 CCACTAGCCAACCACTTTTTGCCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGGC 3060

Qy   3061 agctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaag 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 AGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGACGAAG 3120

Qy   3121 ccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccagt 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 CCCTACACAGGGACCTGGCCAACTTCAGGATCCAACACCCTCAGGTGACCCTCCTCCAGT 3180

Qy   3181 acgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaagg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 ACGTGGATGACCTGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGTACGAAGG 3240

Qy   3241 cactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagattt 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 CACTACTGCTGGAATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATTT 3300

Qy   3301 gcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacgg 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 GCAGGAGAGAGGTAACATACTTGGGGTACAGTTTGCGGGGCGGGCAGCGATGGCTGACGG 3360

Qy   3361 aggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgagag 3420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3361 AGGCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAGTGAGAG 3420

Qy   3421 agtttttggggacagctggattttgcagactgtggatcccggggtttgcgaccttagcag 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 AGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTTAGCAG 3480

Qy   3481 ccccactctacccgctaaccaaagaaaaaggggattctcctgggctcctgagcaccaga 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CCCCACTCTACCCGCTAACCAAAGAAAAAGGGGGATTCTCCTGGGCTCCTGAGCACCAGA 3540

Qy   3541 aggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctccctgacg 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 AGGCATTTGATGCTATCAAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCCTGACG 3600

Qy   3601 taactaaacccttttaccctttatgtggatgagcgtaagggagtagcccgaggagttttaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 TAACTAAACCCTTTACCCTTTATGTGGATGAGCGTAAGGGAGTAGCCCGAGGAGTTTTAA 3660

Qy   3661 cccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagcttgatcctg 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 CCCAAACCCTAGGACCATGGAGGAGACCTGTTGCCTACCTGTCAAAGAAGCTTGATCCTG 3720

Qy   3721 tagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccatactggtcaagg 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 TAGCCAGTGGTTGGCCCGTATGTCTGAAGGCTATCGCAGCTGTGGCCATACTGGTCAAGG 3780

Qy   3781 acgctgacaaattgactttgggacagaatataactgtaatagccccccatgcattggaga 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 ACGCTGACAAATTGACTTTGGGACAGAATATAACTGTAATAGCCCCCCATGCATTGGAGA 3840

Qy   3841 acatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacccactatcaaa 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 ACATCGTTCGGCAGCCCCCAGACCGATGGATGACCAACGCCCGCATGACCCACTATCAAA 3900

Qy   3901 gcctgcttctcacagagagggtcactttcgctccaccagccgctctcaaccctgccactc 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 GCCTGCTTCTCACAGAGAGGGTCACTTTCGCTCCACCAGCCGCTCTCAACCCTGCCACTC 3960

Qy   3961 ttctgcctgaagagactgatgaaccagtgactcatgattgccatcaactattgattgagg 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 TTCTGCCTGAAGAGACTGATGAACCAGTGACTCATGATTGCCATCAACTATTGATTGAGG 4020

Qy   4021 agactggggtccgcaaggaccttacagacataccgctgactggagaagtgctaacctggt 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 AGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGGAGAAGTGCTAACCTGGT 4080

Qy   4081 tcactgacggaagcagctatgtggtggaaggtaagaggatggctggggcggcagtggtgg 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 TCACTGACGGAAGCAGCTATGTGGTGGAAGGTAAGAGGATGGCTGGGGCGGCAGTGGTGG 4140

Qy   4141 acgggacccgcacgatctgggccagcagcctgccggaaggaacttcagcgcaaaaggctg 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 ACGGGACCCGCACGATCTGGGCCAGCAGCCTGCCGGAAGGAACTTCAGCGCAAAAGGCTG 4200

Qy   4201 agctcatggccctcacgcaagctttgcggctggccgaagggaaatccataaacatttata 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 AGCTCATGGCCCTCACGCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAAACATTTATA 4260

Qy   4261 cggacagcaggtatgcctttgcgactgcacacgtacacggggccatctataaacaaaggg 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 CGGACAGCAGGTATGCCTTTGCGACTGCACACGTACACGGGGCCATCTATAAACAAAGGG 4320

Qy   4321 ggttgcttacctcagcagggagggaaataaagaacaaagaggaaattctaagcctattag 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAGGAAATTCTAAGCCTATTAG 4380

Qy   4381 aagccttacatttgccaaaaaggctagctattatacactgtcctggacatcagaaagcca 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 AAGCCTTACATTTGCCAAAAAGGCTAGCTATTATACACTGTCCTGGACATCAGAAAGCCA 4440

Qy   4441 aagatctcatatctagagggaaccagatggctgaccgggttgccaagcaggcagcccagg 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 AAGATCTCATATCTAGAGGGAACCAGATGGCTGACCGGGTTGCCAAGCAGGCAGCCCAGG 4500

Qy   4501 ctgttaaccttctgcctataatagaaacgcccaaagccccagaacccagacgacagtaca 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 CTGTTAACCTTCTGCCTATAATAGAAACGCCCAAAGCCCCAGAACCCAGACGACAGTACA 4560

Qy   4561 ccctagaagactggcaagagataaaaaagatagaccagttctctgagactccggagggga 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 CCCTAGAAGACTGGCAAGAGATAAAAAAGATAGACCAGTTCTCTGAGACTCCGGAGGGGA 4620

Qy   4621 cctgctatacctcatatgggaaggaaatcctgccccacaaagaagggttagaatatgtcc 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 CCTGCTATACCTCATATGGGAAGGAAATCCTGCCCCACAAAGAAGGGTTAGAATATGTCC 4680

Qy   4681 aacagatacatcgtctaacccacctaggaactaaacacctgcagcagttggtcagaacat 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 AACAGATACATCGTCTAACCCACCTAGGAACTAAACACCTGCAGCAGTTGGTCAGAACAT 4740

Qy   4741 ccccttatcatgttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgc 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 CCCCTTATCATGTTCTGAGGCTACCAGGAGTGGCTGACTCGGTGGTCAAACATTGTGTGC 4800

Qy   4801 cctgccagctggttaatgctaatccttccagaatacctccaggaaagagactaaggggaa 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 CCTGCCAGCTGGTTAATGCTAATCCTTCCAGAATACCTCCAGGAAAGAGACTAAGGGGAA 4860

Qy   4861 gccacccaggcgctcactgggaagtggacttcactgaggtaaagccggctaaatacggaa 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 GCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAATACGGAA 4920

Qy   4921 acaaatatctattggtttttgtagacacctttcaggatgggtagaggcttatcctacta 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 ACAAATATCTATTGGTTTTTGTAGACACCTTTCAGGATGGGTAGAGGCTTATCCTACTA 4980

Qy   4981 agaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttccaagatttg 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 AGAAAGAGACTTCAACCGTGGTGGCTAAGAAAATACTGGAGGAAATTTTTCCAAGATTTG 5040

Qy   5041 gaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcagg 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 GAATACCTAAGGTAATAGGGTCAGACAATGGTCCAGCTTTCGTTGCCCAGGTAAGTCAGG 5100

Qy   5101 gactggccaagatattggggattgattggaaactgcattgtgcatacagaccccaaagct 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 GACTGGCCAAGATATTGGGGATTGATTGGAAACTGCATTGTGCATACAGACCCCAAAGCT 5160

Qy   5161 caggacaggtagagaggatgaatagaaccattaaagagacccttaccaaattgaccacag 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 CAGGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACCCTTACCAAATTGACCACAG 5220

Qy   5221 agactggcattaatgattggatggctctcctgccctttgtgctttttagggtgaggaaca 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 AGACTGGCATTAATGATTGGATGGCTCTCCTGCCCTTTGTGCTTTTTAGGGTGAGGAACA 5280

Qy   5281 cccctggacagtttgggctgacccccctatgaattgctctacgggggaccccccccgttgg 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 CCCCTGGACAGTTTGGGCTGACCCCCTATGAATTGCTCTACGGGGGACCCCCCCCGTTGG 5340

Qy   5341 cagaaattgcctttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggc 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 CAGAAATTGCCTTTGCACATAGTGCTGATGTGCTGCTTTCCCAGCCTTTGTTCTCTAGGC 5400

Qy   5401 tcaaggcgctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcag 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 TCAAGGCGCTCGAGTGGGTGAGGCAGCGAGCGTGGAAGCAGCTCCGGGAGGCCTACTCAG 5460

Qy   5461 gaggagacttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgcc 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 GAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTTGGAGATTCAGTCTATGTTAGACGCC 5520

Qy   5521 accgtgcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacac 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 ACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGACCTTATCTCGTACTTTTGACCACAC 5580
```

```
Qy   5581 caacggctgtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccgg 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 CAACGGCTGTGAAAGTCGAAGGAATCCCCACCTGGATCCATGCATCCCACGTTAAGCCGG 5640

Qy   5641 cgccacctcccgactcggggtggagagccgaaaagactgagaatcccctaagcttcgcc 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 CGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGACTGAGAATCCCCTTAAGCTTCGCC 5700

Qy   5701 tccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttata 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701 TCCATCGCCTGGTTCCTTACTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTATA 5760

Qy   5761 gacagctcgaaccccccatagacctttatccctacctggctgattattgaccctgatacg 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 GACAGCTCGAACCCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGATACG 5820

Qy   5821 ggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaactg 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 GGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGTGGCCTGAACTG 5880

Qy   5881 catttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtccgt 5940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5881 CATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGCACACCTCCCAACCTAGTCCGT 5940

Qy   5941 agttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctggg 6000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5941 AGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGAGAAATACTGTGGGGGTTCTGGG 6000

Qy   6001 gaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggccg 6060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6001 GAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGCCG 6060

Qy   6061 atctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtacaaa 6120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6061 ATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACAAA 6120

Qy   6121 atgatgaaactatataaagataagagctgctccccatcagacttagattatctaaagata 6180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6121 ATGATGAAACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTATCTAAAGATA 6180

Qy   6181 agtttcactgaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagctgggg 6240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6181 AGTTTCACTGAAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTATGAGCTGGGG 6240

Qy   6241 aatagtttttttattatatggcgggggagcagggtccacttttaaccattcgccttaggata 6300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6241 AATAGTTTTTTATTATATGGCGGGGGAGCAGGGTCCACTTTAACCATTCGCCTTAGGATA 6300

Qy   6301 gagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaacagggg 6360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6301 GAGACGGGGACAGAACCCCCTGTGGCAATGGGACCCGATAAAGTACTGGCTGAACAGGGG 6360

Qy   6361 cccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgcggcct 6420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6361 CCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGCGGCCT 6420

Qy   6421 gacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgcctaga 6480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6421 GACATAACACAGCCGCCTAGCAACAGTACCACTGGATTGATTCCTACCAACACGCCTAGA 6480

Qy   6481 aactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagggagct 6540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6481 AACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGGGAGCT 6540

Qy   6541 ttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtctatcc 6600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6541 TTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTTGTCTATCC 6600

Qy   6601 tcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagag 6660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6601 TCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAAGAAAATTCAATGTGACCAAAGAG 6660

Qy   6661 catagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagtttccggg 6720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6661 CATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTTCCGGG 6720

Qy   6721 aaggggacatgcataggaaaagctcccccatcccaccaacaccctttgctatagtactgtg 6780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6721 AAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAGTACTGTG 6780

Qy   6781 gtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggca 6840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6781 GTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGTGGGCA 6840

Qy   6841 tgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttc 6900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6841 TGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAGATTTC 6900

Qy   6901 tgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgat 6960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6901 TGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCCTTGAT 6960

Qy   6961 gaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctagctgta 7020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6961 GAATATGACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAGCTGTA 7020

Qy   7021 atgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatcacagga 7080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7021 ATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTGCCCTGATCACAGGA 7080

Qy   7081 ccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatctccga 7140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7081 CCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGATCTCCGA 7140

Qy   7141 gccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaagtggtt 7200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7141 GCCTTAAAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAGTGGTT 7200

Qy   7201 ctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtgcagcc 7260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7201 CTACAGAACCGGAGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTGCAGCC 7260

Qy   7261 ttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactccatgaac 7320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7261 TTAAAAGAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCATGAAC 7320

Qy   7321 aagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggtggttt 7380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7321 AAGCTTAGAAAAAAGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTT 7380

Qy   7381 gaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggccc 7440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7381 GAAGGATGGTTCAACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCC 7440

Qy   7441 ctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgcc 7500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7441 CTAGTAGTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCC 7500

Qy   7501 tttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggc 7560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7501 TTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGC 7560

Qy   7561 cttctgagccaaggagaaactgacctctagccttcccagttctaagattagaactattaa 7620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7561 CTTCTGAGCCAAGGAGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAA 7620

Qy   7621 caagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccagg 7680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7621 CAAGACAAGAAGTGGGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGG 7680

Qy   7681 aagttaataaaaagctctaaatgccccgaattacagaccctgctggctgccagtaaata 7740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7681 AAGTTAATAAAAAGCTCTAAATGCCCCCGAATTACAGACCCTGCTGGCTGCCAGTAAATA 7740

Qy   7741 ggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaagataac 7800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    7741 GGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAAC 7800

Qy    7801 aggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccatagaag 7860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7801 AGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACCATAGAAG 7860

Qy    7861 aattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactctgccc 7920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7861 AATTGATTACACATTGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTCTGCCC 7920

Qy    7921 aggagcccacgcagatgcggacctccggagctattttaaaatgattggtccacggagcgc 7980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7921 AGGAGCCCACGCAGATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGGAGCGC 7980

Qy    7981 gggctctcgatattttaaaatgattggtccatggagcgcgggctctcgatattttaaaat 8040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7981 GGGCTCTCGATATTTTAAAATGATTGGTCCATGGAGCGCGGGCTCTCGATATTTTAAAAT 8040

Qy    8041 gattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgattccgca 8100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8041 GATTGGTTTGTGACGCACAGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCGCA 8100

Qy    8101 ctcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgcttgg 8160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8101 CTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCTTGG 8160

Qy    8161 aataaaaatcctcttgctgtttgcatcaaaaaaaaaaaaaaaaaaaaa 8209
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    8161 AATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAAAAAAAAAAAAAA 8209


RESULT   2
A66552
LOCUS       A66552       8196 bp    DNA              PAT       29-MAR-1999
DEFINITION  Sequence 2 from Patent WO9740167.
ACCESSION   A66552
VERSION     A66552.1  GI:4538105
KEYWORDS    .
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 8196)
  AUTHORS   Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
  TITLE     PORCINE RETROVIRUS
  JOURNAL   Patent: WO 9740167-A 2 30-OCT-1997;
            Q ONE BIOTECH LTD (GB)
FEATURES             Location/Qualifiers
     source          1. .8196
                     /organism="porcine endogenous retrovirus"
                     /db_xref="taxon:61673"
BASE COUNT     2165 a    2061 c    2146 g    1820 t       4 others
ORIGIN


  Query Match             94.4%;  Score 7751;  DB 6;  Length 8196;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 8191;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy      13 tgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatcaagaccgcttct 72
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATCAAGACCGCTTCT 60

Qy      73 cgtgagtgatttggggtgtcgcctcttccgagcccggacgagggggattgttcttttact 132
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      61 CGTGAGTGATTTGGGGTGTCGCCTCTTCCGAKCCCGGACGAGGGGGATTGTTCTTTTACT 120

Qy     133 ggcctttcatttggtgcgttggccgggaaatcctgcgaccacccccttacacccgagaacc 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGCCTTTCATTTGGTGCGTTGGCCGGGAAATCCTGCGACCACCCCTTACACCCGAGAACC 180

Qy     193 gacttggaggtaaagggatccccctttggaacgtgtgtgtgtgtcggccggcgtctctgtt 252
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     181 GACTTGGAGGTAAAGGGATCCCCTTTGGAACGTGTGTGTGTGTCGGCCGGCGTCTCTGTT 240

Qy     253 ctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcctctcagaccgtaa 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTGAGTGTCTGTTTTCGGTGATGCGCGCTTTCGGTTTGCAGCTGTCCTCTCAGACCGTAA 300

Qy     313 ggactggaggactgtgatcagcagacgtgctaggaggatcacaggctgccaccctggggg 372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGACTGGAGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAGGCTGCCACCCTGGGGG 360

Qy     373 acgccccgggaggtggggagagccagggacgcctggtggtctcctactgtcggtcagagg 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ACGCCCCGGGAGGTGGGGAGAGCCAGGGACGCCTGGTGGTCTCCTACTGTCGGTCAGAGG 420

Qy     433 accgagttctgttgttgaagcgaaagcttccccctccgcggccgtccgactcttttgcct 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCGACTCTTTTGCCT 480

Qy     493 gcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggtttctgtctcgtgtg 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GCTTGTGGAAGACGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGTTTCTGTCTCGTGTG 540

Qy     553 tctttgtcttgtgcgtccttgtctacagttttaatatgggacagacagtgactacccccc 612
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TCTTTGTCTTGTGCGTCCTTGTCTACAGTTTTAATATGGGACAGACAGTGACTACCCCCC 600

Qy     613 ttagtttgactctcgaccattggactgaagttagatccagggctcataatttgtcagttc 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTAGTTTGACTCTCGACCATTGGACTGAAGTTAGATCCAGGGCTCATAATTTGTCAGTTC 660

Qy     673 aggttaagaagggaccttggcagactttctgtgcctctgaatggccaacattcgatgttg 732
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGCCTCTGAATGGCCAACATTCGATGTTG 720

Qy     733 gatggccatcagaggggacctttaattctgaaattatcctggctgttaaggcaatcattt 792
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GATGGCCATCAGAGGGGACCTTTAATTCTGAAATTATCCTGGCTGTTAAGGCAATCATTT 780

Qy     793 ttcagactggacccggctctcatcctgatcaggagccctatatccttacgtggcaagatt 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TTCAGACTGGACCCGGCTCTCATCCTGATCAGGAGCCCTATATCCTTACGTGGCAAGATT 840

Qy     853 tggcagaagatcctccgccatgggttaaaccatggctaaataaaccaagaaagccaggtc 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TGGCAGAAGATCCTCCGCCATGGGTTAAACCATGGCTAAATAAACCAAGAAAGCCAGGTC 900

Qy     913 cccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaagtcgagccctctt 972
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CCCGAATCCTGGCTCTTGGAGAGAAAAACAAACACTCGGCCGAAAAAGTCGAGCCCTCTT 960

Qy     973 cctcgtatctaccccgagatcgaggagccgccgacttggccggaaccccaacctgttccc 1032
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CCTCGTATCTACCCCGAGATCGAGGAGCCGCCGACTTGGCCGGAACCCCAACCTGTTCCC 1020

Qy    1033 ccacccccttatccagcacagggtgctgtgaggggacctctgcccctcctggagctccgg 1092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCACCCCCTTATCCAGCACAGGGTGCTGTGAGGGGACCTCTGCCCCTCCTGGAGCTCCGG 1080

Qy    1093 tggtggagggacctgctgccgggactcggagccggagaggcgccaccccggagcggacag 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TGGTGGAGGGACCTGCTGCCGGGACTCGGAGCCGGAGAGGCGCCACCCCGGAGCGGACAG 1140

Qy    1153 acgagatcgcgatattaccgctgcgcacctatggccctcccatgccagggggccaattgc 1212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ACGAGATCGCGATATTACCGCTGCGCACCTATGGCCCTCCCATGCCAGGGGGCCAATTGC 1200

Qy    1213 agcccctccagtattggccctttttcttctgcagatctctataattggaaaactaaccatc 1272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 AGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGGAAAACTAACCATC 1260

Qy    1273 ccccttttctcggaggatccccaacgcctcacggggttggtggagtcccttatgttctctc 1332
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGAGTCCCTTATGTTCTCTC 1320
```

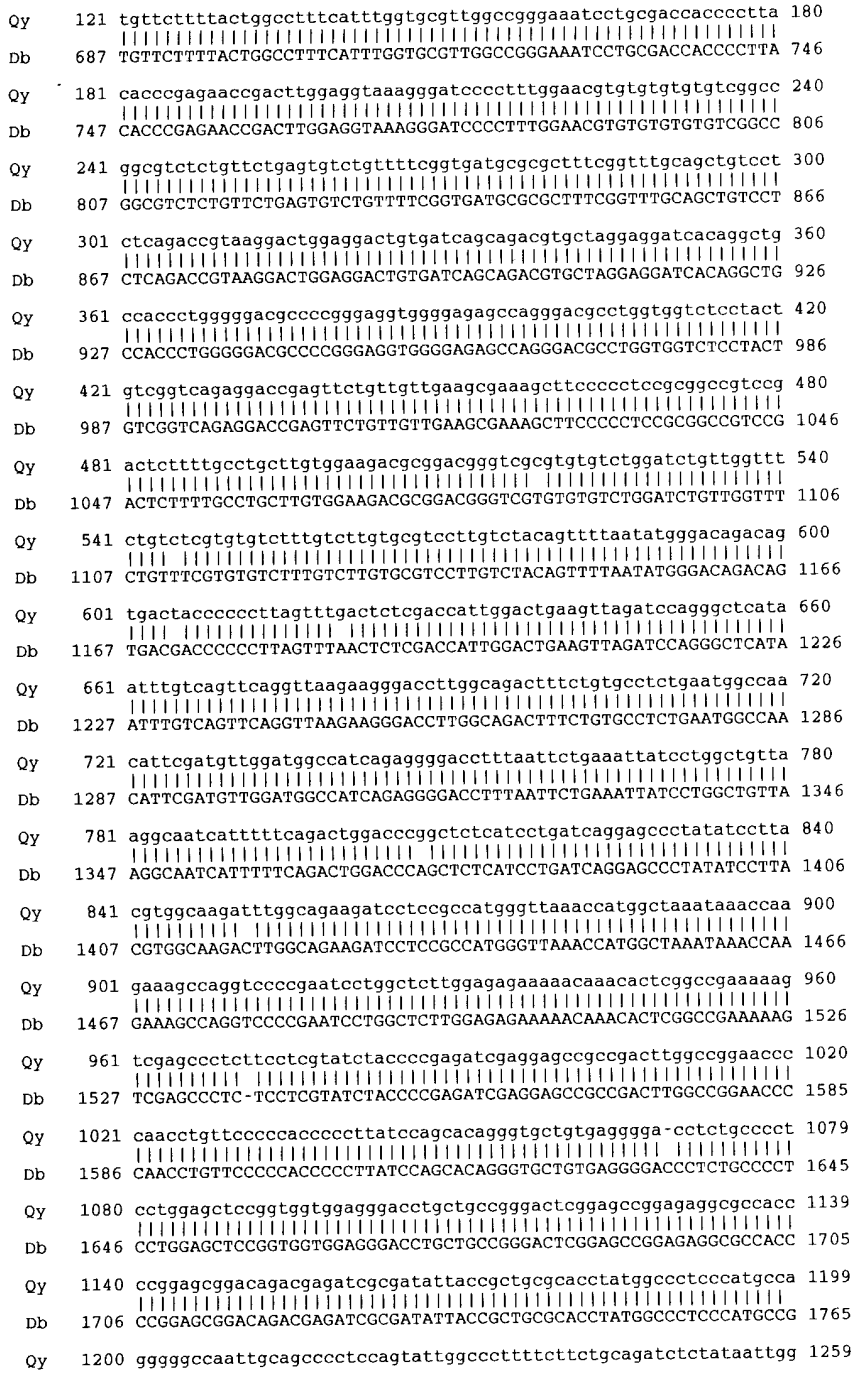
```
Qy   1333 accagcctacttgggatgattgtcaacagctgctgcagacactcttcacaaccgaggagc 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTCACAACCGAGGAGC 1380

Qy   1393 gagagagaattctgttagaggctagaaaaaatgttcctggggccgacgggcgacccacgc 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GAGAGAGAATTCTGTTAGAGGCTAGAAAAAATGTTCCTGGGGCCGACGGGCGACCCACGC 1440

Qy   1453 agttgcaaaatgagattgacatgggatttcccttgactcgcccccggttgggactacaaca 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AGTTGCAAAATGAGATTGACATGGGATTTCCCTTGACTCGCCCCGGTTGGGACTACAACA 1500

Qy   1513 cggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcgggtctccggg 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCTGGTGGCGGGTCTCCGGG 1560

Qy   1573 gcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcagggaccgaacg 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAGGTGATGCAGGGACCGAACG 1620

Qy   1633 aacctccctcggtatttcttgagaggctcatggaagccttcaggcggttcacccctttg 1692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AACCTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGGTTCACCCCTTTTG 1680

Qy   1693 atcctacctcagaggcccagaaagcctcagtggccctggccttcattgggcagtcggctc 1752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ATCCTACCTCAGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTCATTGGGCAGTCGGCTC 1740

Qy   1753 tggatatcaggaagaaacttcagagactggaagggttacaggaggctgagttacgtgatc 1812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TGGATATCAGGAAGAAACTTCAGAGACTGGAAGGGTTACAGGAGGCTGAGTTACGTGATC 1800

Qy   1813 tagtgagagaggcagagaaggtgtattacagaagggagacagaagaggagaaggaacaga 1872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TAGTGAGAGAGGCAGAGAAGGTGTATTACAGAAGGGAGACAGAAGAGGAGAAGGAACAGA 1860

Qy   1873 gaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacggcaagagaagaatt 1932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GAAAAGAAAAGGAGAGAGAAGAAAGGGAGGAAAGACGTGATAGACGGCAAGAGAAGAATT 1920

Qy   1933 tgactaagatcttggccgcagtggttgaagggaagagcagcagggagagagagagagatt 1992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TGACTAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAGAGAGAGAGAGATT 1980

Qy   1993 ttaggaaaattaggtcaggccctagacagtcagggaacctgggcaataggaccccactcg 2052
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAATAGGACCCCACTCG 2040

Qy   2053 acaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaaggaactgccccaaga 2112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 ACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGGACACTGGGCAAGGAACTGCCCCAAGA 2100

Qy   2113 agggaaacaaaggaccgaaggtcctagctctagaagaagataaagattaggggagacggg 2172
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||| ||
Db   2101 AGGGAAACAAAGGACCGAA-GTCCTAGCTCTAGAAGAAGATAAAGATTAGGGGAGAC-GG 2158

Qy   2173 gttcggacccccctccccgagcccagggtaactttgaaggtggaggggcaaccagttgagt 2232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2159 GTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGAGGGGCAACCAGTTGAGT 2218

Qy   2233 tcctggttgataccggagcggagcattcagtgctgctacaaccattaggaaaactaaaag 2292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2219 TCCTGGTTGATACCGGAGCGGAGCATTCAGTGCTGCTACAACCATTAGGAAAACTAAAAG 2278

Qy   2293 aaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatccatggactacccgaa 2352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2279 AAAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACGGCAGTATCCATGGACTACCCGAA 2338

Qy   2353 gaaccgttgacttgggagtgggacgggtaacccactcgtttctggtcatccctgagtgcc 2412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2339 GAACCGTTGACTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTCATCCCTGAGTGCC 2398

Qy   2413 cagtacccccttctaggtagagacttactgaccaagatgggagctcaaatttcttttgaac 2472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2399 CAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAAATTTCTTTTGAAC 2458

Qy   2473 aaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttgaccctccaattag 2532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2459 AAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCACTGTGTTGACCCTCCAATTAG 2518

Qy   2533 atgatgaatatcgactatattctccccaagtaaagcctgatcaagatatacagtcctggt 2592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2519 ATGATGAATATCGACTATATTCTCCCCAAGTAAAGCCTGATCAAGATATACAGTCCTGGT 2578

Qy   2593 tggagcagtttccccaagcctgggcagaaaccgcagggatgggtttggcaaagcaagttc 2652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2579 TGGAGCAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGGTTTGGCAAAGCAAGTTC 2638

Qy   2653 ccccacaggttattcaactgaaggccagtgctacaccagtatcagtcagacagtacccct 2712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2639 CCCCACAGGTTATTCAACTGAAGGCCAGTGCTACACCAGTATCAGTCAGACAGTACCCCT 2698

Qy   2713 tgagtagagaggctcgagaaggaatttggccgcatgttcaaagattaatccaacagggca 2772
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2699 TGAGTAGAGAGGCTCGAGAAGGAATTTGGCCGCATGTTCAAAGATTAATCCAACAGGGCA 2758

Qy   2773 tcctagttcctgtccaatcccccttggaatactccccctgctaccggttaggaagcctggga 2832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2759 TCCTAGTTCCTGTCCAATCCCCTTGGAATACTCCCCTGCTACCGGTTAGGAAGCCTGGGA 2818

Qy   2833 ccaatgattatcgaccagtacaggacttgagagaggtcaataaaagggtgcaggacatac 2892
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2819 CCAATGATTATCGACCAGTACAGGACTTGAGAGAGGTCAATAAAAGGGTGCAGGACATAC 2878

Qy   2893 acccaacggtcccgaacccttataacctcttgagcgccctcccgcctgaacggaactggt 2952
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2879 ACCCAACGGTCCCGAACCCTTATAACCTCTTGAGCGCCCTCCCGCCTGAACGGAACTGGT 2938

Qy   2953 acacagtattggacttaaaagatgccttcttctgcctgagattacaccccactagccaac 3012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2939 ACACAGTATTGGACTTAAAAGATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAAC 2998

Qy   3013 cacttttttgccttcgaatggagagatccaggtacgggaagaaccgggcagctcacctgga 3072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2999 CACTTTTTGCCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGA 3058

Qy   3073 cccgactgccccaagggttcaagaactccccgaccatctttgacgaagccctacacaggg 3132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3059 CCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGACGAAGCCCTACACAGGG 3118

Qy   3133 acctggccaacttcaggatccaacaccctcaggtgaccctcctccagtacgtggatgacc 3192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3119 ACCTGGCCAACTTCAGGATCCAACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACC 3178

Qy   3193 tgcttctggcgggagccaccaaacaggactgcttagaaggtacgaaggcactactgctgg 3252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3179 TGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGTACGAAGGCACTACTGCTGG 3238

Qy   3253 aattgtctgacctaggctacagagcctctgctaagaaggcccagatttgcaggagagagg 3312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3239 AATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGG 3298

Qy   3313 taacatacttggggtacagtttgcggggcgggcagcgatggctgacggaggcacggaaga 3372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3299 TAACATACTTGGGGTACAGTTTGCGGGGCGGGCAGCGATGGCTGACGGAGGCACGGAAGA 3358

Qy   3373 aaactgtagtccagataccggccccaaccacagccaaacaagtgagagagtttttgggga 3432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3359 AAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAGTGAGAGAGTTTTTGGGGA 3418

Qy   3433 cagctggattttgcagactgtggatcccggggtttgcgaccttagcagccccactctacc 3492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3419 CAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACC 3478

Qy   3493 cgctaaccaaagaaaaaggggggattctcctgggctcctgagcaccagaaggcatttgatg 3552
```

```
Db   3479 CGCTAACCAAAGAAAAAGGGGGATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATG 3538

Qy   3553 ctatcaaaaaggccctgctgagcgcacctgctctggccctccctgacgtaactaaaccct 3612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3539 CTATCAAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCT 3598

Qy   3613 ttacccttttatgtggatgagcgtaagggagtagcccgaggagttttaacccaaaccctag 3672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3599 TTACCCTTTATGTGGATGAGCGTAAGGGAGTAGCCCGAGGAGTTTTAACCCAAACCCTAG 3658

Qy   3673 gaccatggaggagacctgttgcctacctgtcaaagaagcttgatcctgtagccagtggtt 3732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3659 GACCATGGAGGAGACCTGTTGCCTACCTGTCAAAGAAGCTTGATCCTGTAGCCAGTGGTT 3718

Qy   3733 ggcccgtatgtctgaaggctatcgcagctgtggccatactggtcaaggacgctgacaaat 3792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3719 GGCCCGTATGTCTGAAGGCTATCGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAAT 3778

Qy   3793 tgactttgggacagaatataactgtaatagccccccatgcattggagaacatcgttcggc 3852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3779 TGACTTTGGGACAGAATATAACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGC 3838

Qy   3853 agcccccagaccgatggatgaccaacgcccgcatgacccactatcaaagcctgcttctca 3912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3839 AGCCCCCAGACCGATGGATGACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCA 3898

Qy   3913 cagagagggtcactttcgctccaccagccgctctcaaccctgccactcttctgcctgaag 3972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3899 CAGAGAGGGTCACTTTCGCTCCACCAGCCGCTCTCAACCCTGCCACTCTTCTGCCTGAAG 3958

Qy   3973 agactgatgaaccagtgactcatgattgccatcaactattgattgaggagactggggtcc 4032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3959 AGACTGATGAACCAGTGACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCC 4018

Qy   4033 gcaaggaccttacagacataccgctgactggagaagtgctaacctggttcactgacggaa 4092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4019 GCAAGGACCTTACAGACATACCGCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAA 4078

Qy   4093 gcagctatgtggtggaaggtaagaggatggctggggcggcagtggtggacgggacccgca 4152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4079 GCAGCTATGTGGTGGAAGGTAAGAGGATGGCTGGGGCGGCAGTGGTGGACGGGACCCGCA 4138

Qy   4153 cgatctgggccagcagcctgccggaaggaacttcagcgcaaaaggctgagctcatggccc 4212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4139 CGATCTGGGCCAGCAGCCTGCCGGAAGGAACTTCAGCGCAAAAGGCTGAGCTCATGGCCC 4198

Qy   4213 tcacgcaagctttgcggctggccgaagggaaatccataaacatttatacggacagcaggt 4272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4199 TCACGCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAAACATTTATACGGACAGCAGGT 4258

Qy   4273 atgcctttgcgactgcacacgtacacggggccatctataaacaaaggggttgcttacct 4332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4259 ATGCCTTTGCGACTGCACACGTACACGGGGCCATCTATAAACAAAGGGGGTTGCTTACCT 4318

Qy   4333 cagcagggagggaaataaagaacaaagaggaaattctaagcctattagaagccttacatt 4392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4319 CAGCAGGGAGGGAAATAAAGAACAAAGAGGAAATTCTAAGCCTATTAGAAGCCTTACATT 4378

Qy   4393 tgccaaaaaggctagctattatacactgtcctggacatcagaaagccaaagatctcatat 4452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4379 TGCCAAAAAGGCTAGCTATTATACACTGTCCTGGACATCAGAAAGCCAAAGATCTCATAT 4438

Qy   4453 ctagagggaaccagatggctgaccgggttgccaagcaggcagcccaggctgttaaccttc 4512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4439 CTAGAGGGAACCAGATGGCTGACCGGGTTGCCAAGCAGGCAGCCCAGGCTGTTAACCTTC 4498

Qy   4513 tgcctataatagaaacgcccaaagccccagaacccagacgacagtacaccctagaagact 4572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4499 TGCCTATAATAGAAACGCCCAAAGCCCCAGAACCCAGACGACAGTACACCCTAGAAGACT 4558

Qy   4573 ggcaagagataaaaaagatagaccagttctctgagactccggaggggacctgctatacct 4632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

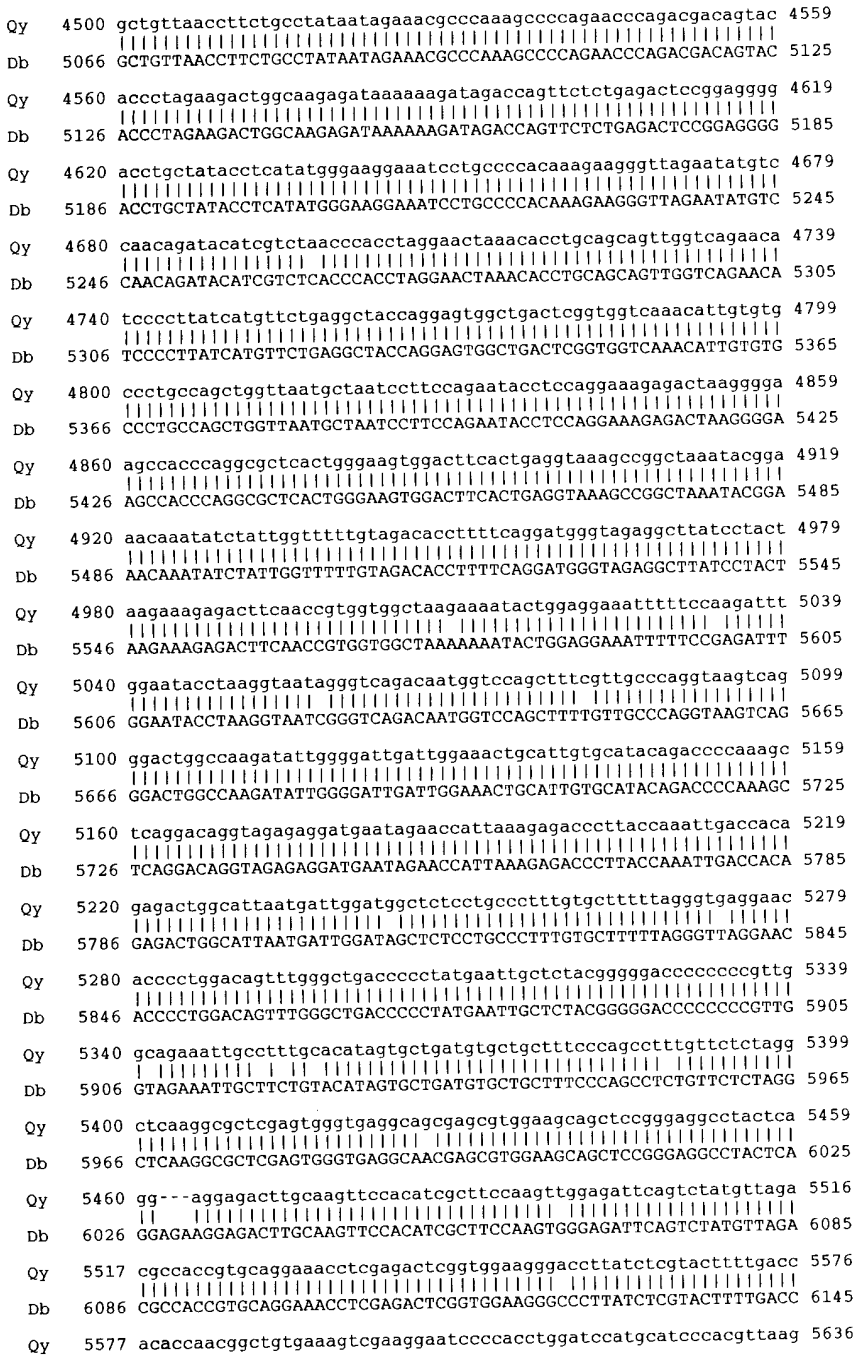
```
Db   4559 GGCAAGAGATAAAAAAGATAGACCAGTTCTCTGAGACTCCGGAGGGGACCTGCTATACCT 4618

Qy   4633 catatgggaaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatc 4692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4619 CATATGGGAAGGAAATCCTGCCCCACAAAGAAGGGTTAGAATATGTCCAACAGATACATC 4678

Qy   4693 gtctaacccacctaggaactaaacacctgcagcagttggtcagaacatccccttatcatg 4752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4679 GTCTAACCCACCTAGGAACTAAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATG 4738

Qy   4753 ttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctgg 4812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4739 TTCTGAGGCTACCAGGAGTGGCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGG 4798

Qy   4813 ttaatgctaatccttccagaatacctccaggaaagagactaaggggaagccacccaggcg 4872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4799 TTAATGCTAATCCTTCCAGAATACCTCCAGGAAAGAGACTAAGGGGAAGCCACCCAGGCG 4858

Qy   4873 ctcactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatctat 4932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4859 CTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAATACGGAAACAAATATCTAT 4918

Qy   4933 tggtttttgtagacacctttcaggatgggtagaggcttatcctactaagaaagagactt 4992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4919 TGGTTTTTGTAGACACCTTTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTT 4978

Qy   4993 caaccgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaagg 5052
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4979 CAACCGTGGTGGCTAAGAAAATACTGGAGGAAATTTTTCCAAGATTTGGAATACCTAAGG 5038

Qy   5053 taatagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaaga 5112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5039 TAATAGGGTCAGACAATGGTCCAGCTTTCGTTGCCCAGGTAAGTCAGGGACTGGCCAAGA 5098

Qy   5113 tattggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggtag 5172
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5099 TATTGGGGATTGATTGGAAACTGCATTGTGCATACAGACCCCAAAGCTCAGGACAGGTAG 5158

Qy   5173 agaggatgaatagaaccattaaagagacccttaccaaattgaccacagagactggcatta 5232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5159 AGAGGATGAATAGAACCATTAAAGAGACCCTTACCAAATTGACCACAGAGACTGGCATTA 5218

Qy   5233 atgattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacagt 5292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5219 ATGATTGGATGGCTCTCCTGCCCTTTGTGCTTTTTAGGGTGAGGAACACCCCTGGACAGT 5278

Qy   5293 ttgggctgacccccctatgaattgctctacgggggaccccccccgttggcagaaattgcct 5352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5279 TTGGGCTGACCCCCTATGAATTGCTCTACGGGGGACCCCCCCCGTTGGCAGAAATTGCCT 5338

Qy   5353 ttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcg 5412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5339 TTGCACATAGTGCTGATGTGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGCGCTCG 5398

Qy   5413 agtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagacttgc 5472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5399 AGTGGGTGAGGCAGCGAGCGTGGAAGCAGCTCCGGGAGGCCTACTCAGGAGGAGACTTGC 5458

Qy   5473 aagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaa 5532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5459 AAGTTCCACATCGCTTCCAAGTTGGAGATTCAGTCTATGTTAGACGCCACCGTGCAGGAA 5518

Qy   5533 acctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacggctgtga 5592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5519 ACCTCGAGACTCGGTGGAAGGGACCTTATCTCGTACTTTTGACCACACCAACGGCTGTGA 5578

Qy   5593 aagtcgaaggaatccccacctggatccatgcatcccacgttaagccggcgccacctcccg 5652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5579 AAGTCGAAGGAATCCCCACCTGGATCCATGCATCCCACGTTAAGCYGGCGCCACCTCCCG 5638

Qy   5653 actcggggtggagagccgaaaagactgagaatcccctaagcttcgcctccatcgcctgg 5712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5639 ACTCGGGGTGGAGAGCCGAAAAGACTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGG 5698
```

```
Qy   5713 ttccttactctaacaataactccccaggccagtagtaaacgccttatagacagctcgaac 5772
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5699 TTCCTTACTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTATAGACAGCTCGAAC 5758

Qy   5773 ccccatagacctttatcccttacctggctgattattgaccctgatacgggtgtcactgta 5832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5759 CCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGATACGGGTGTCACTGTA 5818

Qy   5833 aatagcactcgaggtgttgctcctagaggcacctggtggcctgaactgcatttctgcctc 5892
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5819 AATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGTGGCCTGAACTGCATTTCTGCCTC 5878

Qy   5893 cgattgattaaccccgctgttaaaagcacacctcccaacctagtccgtagttatgggttc 5952
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   5879 CGATTGATTAACCCCGCTGTTAARAGCACACCTCCCAACCTAGTCCGTAGTTATGGGTTC 5938

Qy   5953 tattgctgcccaggcacagagaaagagaaatactgtgggggttctggggaatccttctgt 6012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5939 TATTGCTGCCCAGGCACAGAGAAAGAGAAATACTGTGGGGGTTCTGGGGAATCCTTCTGT 5998

Qy   6013 aggagatggagctgcgtcacctccaacgatggagactggaaatggccgatctctctccag 6072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5999 AGGAGATGGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGCCGATCTCTCTCCAG 6058

Qy   6073 gaccgggtaaaattctcctttgtcaattccggcccgggcaagtacaaaatgatgaaacta 6132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6059 GACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACAAAATGATGAAACTA 6118

Qy   6133 tataaagataagagctgctccccatcagacttagattatctaaagataagtttcactgaa 6192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6119 TATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTATCTAAAGATAAGTTTCACTGAA 6178

Qy   6193 aggaaaacaggaaaatattcaaaagtggataaatggtatgagctggggaatagttttttta 6252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6179 AGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTA 6238

Qy   6253 ttatatggcgggggagcagggtccactttaaccattcgccttaggatagagacggggaca 6312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6239 TTATATGGCGGGGGAGCAGGGTCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACA 6298

Qy   6313 gaaccccctgtggcaatgggacccgataaagtactggctgaacaggggccccggccctg 6372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6299 GAACCCCCTGTGGCAATGGGACCCGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTG 6358

Qy   6373 gagccaccgcataacttgccggtgccccaattaacctcgctgcggcctgacataacacag 6432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6359 GAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGCGGCCTGACATAACACAG 6418

Qy   6433 ccgcctagcaacagtaccactggattgattcctaccaacacgcctagaaactccccaggt 6492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6419 CCGCCTAGCAACAGTACCACTGGATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGT 6478

Qy   6493 gttcctgttaagacaggacagagactcttcagtctcatccagggagctttccaagccatc 6552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6479 GTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATC 6538

Qy   6553 aactccaccgaccctgatgccacttcttcttgttggctttgtctatcctcagggcctcct 6612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6539 AACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCT 6598

Qy   6613 tattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaa 6672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6599 TATTATGAGGGGATGGCTAAAGAAAGAAAATTCAATGTGACCAAAGAGCATAGAAATCAA 6658

Qy   6673 tgtacatgggggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgc 6732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6659 TGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGC 6718

Qy   6733 ataggaaaagctcccccatcccaccaacacctttgctatagtactgtggtttatgagcag 6792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6719 ATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAG 6778

Qy   6793 gcctcagaaaatcagtatttagtacctggttataacaggtggtgggcatgcaatactggg 6852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6779 GCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGG 6838

Qy   6853 ttaaccccctgtgtttccaccctcagtcttcaaccaatccaaagatttctgtgtcatggtc 6912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6839 TTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTC 6898

Qy   6913 caaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactat 6972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6899 CAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCCTTGATGAATATGACTAT 6958

Qy   6973 cggtataaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggatta 7032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6959 CGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTA 7018

Qy   7033 gggacggccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagcta 7092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7019 GGGACGGCCGTTGGCGTAGGAACAGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTA 7078

Qy   7093 gagaaaggacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggag 7152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7079 GAGAAAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGATCTCCGAGCCTTAAAGGAG 7138

Qy   7153 tctgttagcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccgg 7212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7139 TCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGG 7198

Qy   7213 aggggattagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaa 7272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7199 AGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAA 7258

Qy   7273 tgttgcttctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaa 7332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7259 TGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCATGAACAAGCTTAGAAAA 7318

Qy   7333 aagttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttc 7392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7319 AAGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTC 7378

Qy   7393 aacaggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctg 7452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7379 AACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTG 7438

Qy   7453 ctcctgttacttacagttgggccttgcttaattaataggtttgttgccttttgttagagaa 7512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7439 CTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAA 7498

Qy   7513 cgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaa 7572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7499 CGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAA 7558

Qy   7573 ggagaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaag 7632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7559 GGAGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAG 7618

Qy   7633 tggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaa 7692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7619 TGGGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAA 7678

Qy   7693 agctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtagaaggtca 7752
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   7679 AGCTCTAAATGCCCCCGAATTMCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCA 7738

Qy   7753 cacttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagtt 7812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7739 CACTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTT 7798

Qy   7813 gactaatcgcttatctggattctgtaaaactgactggcaccatagaagaattgattacac 7872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7799 GACTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACCATAGAAGAATTGATTACAC 7858

Qy   7873 attgacagccctagtgacctatctcaactgcaatctgtcactctgcccaggagcccacgc 7932
```

```
Db    7859 ATTGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGC 7918

Qy    7933 agatgcggacctccggagctattttaaaatgattggtccacggagcgcgggctctcgata 7992
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7919 AGATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATA 7978

Qy    7993 ttttaaaatgattggtccatggagcgcgggctctcgatattttaaaatgattggtttgtg 8052
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7979 TTTTAAAATGATTGGTCCATGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTTTGTG 8038

Qy    8053 acgcacaggctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgca 8112
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8039 ACGCACAGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCA 8098

Qy    8113 gtcctctacccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcct 8172
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8099 GTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCT 8158

Qy    8173 cttgctgtttgcatcaaaaaaaaaaaaaaaaaaaaaa 8209
           ||||||||||||||||||||||||||||||||||||
Db    8159 CTTGCTGTTTGCATCAAAAAAAAAAAAAAAAAAAAAA 8195


RESULT   3
PEN133816
LOCUS       PEN133816    8918 bp    DNA             VRL       07-DEC-2000
DEFINITION  Porcine endogenous retrovirus type C proviral gag, pol and env
            genes and LTR (class B, clone 33).
ACCESSION   AJ133816
VERSION     AJ133816.1  GI:6688945
KEYWORDS    env gene; env protein; gag gene; gag protein; pol gene; pol
            protein.
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 8918)
  AUTHORS   Czauderna,F., Fischer,N., Boller,K., Kurth,R. and Tonjes,R.R.
  TITLE     Establishment and characterization of molecular clones of porcine
            endogenous retroviruses replicating on human cells
  JOURNAL   J. Virol. 74 (9), 4028-4038 (2000)
  MEDLINE   20219394
REFERENCE   2  (bases 1 to 8918)
  AUTHORS   Toenjes,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-1999) Toenjes R.R., Medical Biotechnology,
            Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, Hessen, D-63225
            Langen, GERMANY
FEATURES             Location/Qualifiers
     source          1. .8918
                     /organism="porcine endogenous retrovirus"
                     /proviral
                     /strain="type C"
                     /db_xref="taxon:61673"
                     /clone="33"
                     /note="grown in PK15 (ATCC No. CCL-33) cell line"
     LTR             1. .707
     misc_feature    545
                     /note="cap site"
     primer_bind     710. .727
                     /note="gly4"
     gene            1154. .2725
                     /gene="gag"
     mat_peptide     1154. .2725
                     /gene="gag"
                     /function="retroviral capsid"
                     /product="Gag protein"
     CDS             join(1154. .2725,2729. .6313,6189. .8162)
                     /note="viral polypeptide"
                     /codon_start=1
                     /protein_id="CAB65339.1"
                     /db_xref="GI:6688946"
                     /db_xref="SPTREMBL:Q9Q1X5"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPSSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPRIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREEREEERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLTLEEDKDGRRGSDPLPEPRVTLKVEGQPVEFLVDTGAEHSVLLQPLGKL
                     KEKKSWVMGATGQRQYPWTTRRTVDLAVGRVTHSFLVIPECPVPLLGRDLLTKMGAQI
                     SFEQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMG
                     LAKQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPL
                     LPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFF
                     CLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQH
                     PQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYS
                     LRGGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKE
                     KGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPW
                     RRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQ
                     PPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETG
                     VRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAE
                     LMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSL
                     LEALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPIIETPKAPEPR
                     RQYTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEGLEYVQQIHRLTHLGTKHLQ
                     QLVRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRIPPGKRLRGSHPGAHWEVDFTE
                     VKPAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKILEEIFPRFGIPKVIGSDNG
                     PAFVAQVSQGLAKILGIDWKLHCAYRPQSSGQVERMNRTIKETLTKLTTETGINDWIA
                     LLPFVLFRVRNTPGQFGLTPYELLYGGPPPLVEIASVHSADVLLSQPLFSRLKALEWV
                     RQRAWKQLREAYSGEGDLQVPHRFQVGDSVYVRRHRAGNLETRWKGPYLVLLTTPTAV
                     KVEGISTWIHTSHVKLAPPPDSGWRAEKTENPLKLRLHRLVPYSNNNSPGQIHPTLSW
                     RHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSKRLIDSSNPHRPLSLTWLIIDP
                     DTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTPPNLVRSYGFYCCPGTEKEKYC
                     GDSGESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVNSGPGKYKVMKLYKDKSCSPSD
                     LDYLKISFTEKGKQENIQKWINGMSWGIVFYKYGGGAGSTLTIRLRIETGTEPPVAVG
                     PDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQSPSNGTTGLIPTNTPRNSPGVPVKT
                     GQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYYEGMAKEGKFNVTKEHRNQCTW
                     GSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQASENQYLVPGYNRWWACNTGL
                     TPCVSTSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEYDYRYNRPKREPVSLTLAVMLG
                     LGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLRALEESVSNLEESLTSLSEVVL
                     QNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDSMSKLRERLERRRREREADQGW
                     FEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLINRFVAFVRERVSAVQIMVLRQQ
                     YQGLLSQGETDL"
     mat_peptide     2729. .6313
                     /gene="pol"
                     /function="retroviral protease and polymerase"
                     /product="Pol protein"
     gene            2729. .6313
                     /gene="pol"
     mat_peptide     6189. .8162
                     /gene="env"
                     /function="retroviral envelope"
                     /product="Env protein"
     gene            6189. .8162
                     /gene="env"
     LTR             8212. .8918
     polyA_signal    8811. .8816
BASE COUNT      2327 a . 2245 c   2334 g   2012 t
ORIGIN


  Query Match             58.9%;  Score 4833;  DB 14;  Length 8918;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 7963;  Conservative    0;  Mismatches   47;  Indels    8;  Gaps    5;

Qy      1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    567 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 626

Qy     61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    627 AAGACCGCTTCTCGTGAGTGATTTGGGGTGTCGCCTCTTCCGAGCCCGGACGAGGGGGAT 686
```


Qy 121 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctta 180
Db 687 TGTTCTTTTACTGGCCTTTCATTTGGTGCGTTGGCCGGGAAATCCTGCGACCACCCCTTA 746
Qy 181 cacccgagaaccgacttggaggtaaagggatccccctttggaacgtgtgtgtgtgtcggcc 240
Db 747 CACCCGAGAACCGACTTGGAGGTAAAGGGATCCCCTTTGGAACGTGTGTGTGTGTCGGCC 806
Qy 241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
Db 807 GGCGTCTCTGTTCTGAGTGTCTGTTTTCGGTGATGCGCGCTTTCGGTTTGCAGCTGTCCT 866
Qy 301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
Db 867 CTCAGACCGTAAGGACTGGAGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAGGCTG 926
Qy 361 ccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 420
Db 927 CCACCCTGGGGGACGCCCCGGGAGGTGGGGAGAGCCAGGGACGCCTGGTGGTCTCCTACT 986
Qy 421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 480
Db 987 GTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCG 1046
Qy 481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
Db 1047 ACTCTTTTGCCTGCTTGTGGAAGACGCGGACGGGTCGTGTGTGTCTGGATCTGTTGGTTT 1106
Qy 541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
Db 1107 CTGTTTCGTGTGTCTTTGTCTTGTGCGTCCTTGTCTACAGTTTTAATATGGGACAGACAG 1166
Qy 601 tgactacccccctagtttgactctcgaccattggactgaagttagatccagggctcata 660
Db 1167 TGACGACCCCCCTTAGTTTAACTCTCGACCATTGGACTGAAGTTAGATCCAGGGCTCATA 1226
Qy 661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
Db 1227 ATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGCCTCTGAATGGCCAA 1286
Qy 721 cattcgatgttggatggccatcagaggggaccttaattctgaaattatcctggctgtta 780
Db 1287 CATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCTGAAATTATCCTGGCTGTTA 1346
Qy 781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 840
Db 1347 AGGCAATCATTTTTCAGACTGGACCCAGCTCTCATCCTGATCAGGAGCCCTATATCCTTA 1406
Qy 841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
Db 1407 CGTGGCAAGACTTGGCAGAAGATCCTCCGCCATGGGTTAAACCATGGCTAAATAAACCAA 1466
Qy 901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
Db 1467 GAAAGCCAGGTCCCCGAATCCTGGCTCTTGGAGAGAAAAACAAACACTCGGCCGAAAAAG 1526
Qy 961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
Db 1527 TCGAGCCCTC-TCCTCGTATCTACCCCGAGATCGAGGAGCCGCCGACTTGGCCGGAACCC 1585
Qy 1021 caacctgttcccccaccccctttatccagcacagggtgctgtgagggga-cctctgcccct 1079
Db 1586 CAACCTGTTCCCCCACCCCCTTATCCAGCACAGGGTGCTGTGAGGGGACCCTCTGCCCCT 1645
Qy 1080 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1139
Db 1646 CCTGGAGCTCCGGTGGTGGAGGGACCTGCTGCCGGGACTCGGAGCCGGAGAGGCGCCACC 1705
Qy 1140 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1199
Db 1706 CCGGAGCGGACAGACGAGATCGCGATATTACCGCTGCGCACCTATGGCCCTCCCATGCCG 1765
Qy 1200 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1259
Db 1766 GGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGG 1825
Qy 1260 aaaactaaccatccccctttctcggaggatccccaacgcctcacgggggttggtggagtcc 1319
Db 1826 AAAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGAGTCC 1885
Qy 1320 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1379
Db 1886 CTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTC 1945
Qy 1380 acaaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgac 1439
Db 1946 ACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAGAAAAAATGTTCCTGGGGCCGAC 2005
Qy 1440 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1499
Db 2006 GGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGATTTCCCTTGACTCGCCCCGGT 2065
Qy 1500 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1559
Db 2066 TGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCTGGTG 2125
Qy 1560 gcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatg 1619
Db 2126 GCGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAGGTGATG 2185
Qy 1620 cagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcgg 1679
Db 2186 CAGGGACCGAACGAACCTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGG 2245
Qy 1680 ttcaccccttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcatt 1739
Db 2246 TTCACCCCTTTTGATCCTACCTCGGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTCATT 2305
Qy 1740 gggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggct 1799
Db 2306 GGGCAGTCGGCTCTGGATATCAGAAAGAAACTTCAGAGACTGGAAGGGTTACAGGAGGCT 2365
Qy 1800 gagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagag 1859
Db 2366 GAGTTACGTGATCTAGTGAGAGAGGCAGAGAAGGTGTATTACAGAAGGGAGACAGAAGAG 2425
Qy 1860 gagaaggaacagagaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacgg 1919
Db 2426 GAGAAGGAACAGAGAAAAGAAAAGGAGAGAGAAGAAAGGGAGGAAAGACGTGATAGACGG 2485
Qy 1920 caagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggag 1979
Db 2486 CAAGAGAAGAATTTGACTAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAG 2545
Qy 1980 agagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaat 2039
Db 2546 AGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAAT 2605
Qy 2040 aggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagg 2099
Db 2606 AGGACCCCACTCGACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGGACACTGGGCAAGG 2665
Qy 2100 aactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagatgaaagat 2159
Db 2666 AACTGCCCCAAGAAGGGAAACAAAGGACCGAAGGTCCTAACTCTAGAAGAAGATAAAGAT 2725
Qy 2160 taggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtggagggg 2219
Db 2726 TAGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGAGGGG 2785
Qy 2220 caaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccatta 2279
Db 2786 CAACCAGTTGAGTTCCTGGTTGATACCGGAGCGGAGCATTCAGTGCTGCTACAACCATTA 2845
Qy 2280 ggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatcca 2339

```
Db   2846 GGAAAACTAAAAGAAAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACGGCAGTATCCA 2905

Qy   2340 tggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtc 2399
          ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   2906 TGGACTACCCGAAGAACCGTTGACTTGGCAGTGGGACGGGTAACCCACTCGTTTCTGGTC 2965

Qy   2400 atccctgagtgcccagtacccctctaggtagagacttactgaccaagatgggagctcaa 2459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2966 ATCCCTGAGTGCCCAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAA 3025

Qy   2460 atttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttg 2519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3026 ATTTCTTTTGAACAAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCACTGTGTTG 3085

Qy   2520 accctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagat 2579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3086 ACCCTCCAATTAGATGATGAATATCGACTATATTCTCCCCAAGTAAAGCCTGATCAAGAT 3145

Qy   2580 atacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttg 2639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3146 ATACAGTCCTGGTTGGAGCAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGGTTTG 3205

Qy   2640 gcaaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtc 2699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3206 GCAAAGCAAGTTCCCCCACAGGTTATTCAACTGAAGGCCAGTGCTACACCAGTATCAGTC 3265

Qy   2700 agacagtacccctgagtagagaggctcgagaaggaatttggccgcatgttcaaagatta 2759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3266 AGACAGTACCCCTTGAGTAGAGAGGCTCGAGAAGGAATTTGGCCGCATGTTCAAAGATTA 3325

Qy   2760 atccaacagggcatcctagttcctgtccaatccccttggaatactcccctgctaccggtt 2819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3326 ATCCAACAGGGCATCCTAGTTCCTGTCCAATCCCCTTGGAATACTCCCCTGCTACCGGTT 3385

Qy   2820 aggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaagg 2879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3386 AGGAAGCCTGGGACCAATGATTATCGACCAGTACAGGACTTGAGAGAGGTCAATAAAAGG 3445

Qy   2880 gtgcaggacatacacccaacggtcccgaacccttataacctcttgagcgccctcccgcct 2939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3446 GTGCAGGACATACACCCAACGGTCCCGAACCCTTATAACCTCTTGAGCGCCCTCCCGCCT 3505

Qy   2940 gaacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacac 2999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3506 GAACGGAACTGGTACACAGTATTGGACTTAAAAGATGCCTTCTTCTGCCTGAGATTACAC 3565

Qy   3000 cccactagccaaccactttttgccttcgaatggagagatccaggtacgggaagaaccggg 3059
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3566 CCCACTAGCCAACCGCTTTTTGCCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGG 3625

Qy   3060 cagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaa 3119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3626 CAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGACGAA 3685

Qy   3120 gccctacacagggacctggccaacttcaggatccaacaccctcaggtgacccctcctccag 3179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3686 GCCCTACACAGGGACCTGGCCAACTTCAGGATCCAACACCCTCAGGTGACCCTCCTCCAG 3745

Qy   3180 tacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaag 3239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3746 TACGTGGATGACCTGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGTACGAAG 3805

Qy   3240 gcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagatt 3299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3806 GCACTACTGCTGGAATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATT 3865

Qy   3300 tgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacg 3359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3866 TGCAGGAGAGAGGTAACATACTTGGGGTACAGTTTGCGGGGCGGGCAGCGATGGCTGACG 3925

Qy   3360 gaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgaga 3419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3926 GAGGCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAGTGAGA 3985

Qy   3420 gagtttttggggacagctggattttgcagactgtggatcccggggtttgcgaccttagca 3479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3986 GAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTTAGCA 4045

Qy   3480 gccccactctacccgctaaccaaagaaaaaggggattctcctgggctcctgagcaccag 3539
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   4046 GCCCCACTCTACCCGCTAACCAAAGAAAAAGGGGAATTCTCCTGGGCTCCTGAGCACCAG 4105

Qy   3540 aaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctccctgac 3599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4106 AAGGCATTTGATGCTATCAAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCCTGAC 4165

Qy   3600 gtaactaaaccctttaccctttatgtggatgagcgtaagggagtagcccgaggagtttta 3659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4166 GTAACTAAACCCTTTACCCTTTATGTGGATGAGCGTAAGGGAGTAGCCCGAGGAGTTTTA 4225

Qy   3660 acccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagcttgatcct 3719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4226 ACCCAAACCCTAGGACCATGGAGGAGACCTGTTGCCTACCTGTCAAAGAAGCTTGATCCT 4285

Qy   3720 gtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccatactggtcaag 3779
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   4286 GTAGCCAGTGGTTGGCCCGTATGCCTGAAGGCTATCGCAGCTGTGGCCATACTGGTCAAG 4345

Qy   3780 gacgctgacaaattgactttgggacagaatataactgtaatagccccccatgcattggag 3839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4346 GACGCTGACAAATTGACTTTGGGACAGAATATAACTGTAATAGCCCCCCATGCATTGGAG 4405

Qy   3840 aacatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacccactatcaa 3899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4406 AACATCGTTCGGCAGCCCCCAGACCGATGGATGACCAACGCCCGCATGACCCACTATCAA 4465

Qy   3900 agcctgcttctcacagagagggtcactttcgctccaccagccgctctcaaccctgccact 3959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4466 AGCCTGCTTCTCACAGAGAGGGTCACTTTCGCTCCACCAGCCGCTCTCAACCCTGCCACT 4525

Qy   3960 cttctgcctgaagagactgatgaaccagtgactcatgattgccatcaactattgattgag 4019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4526 CTTCTGCCTGAAGAGACTGATGAACCAGTGACTCATGATTGCCATCAACTATTGATTGAG 4585

Qy   4020 gagactggggtccgcaaggaccttacagacataccgctgactggagaagtgctaacctgg 4079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4586 GAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGGAGAAGTGCTAACCTGG 4645

Qy   4080 ttcactgacggaagcagctatgtggtggaaggtaagaggatggctggggcggcagtggtg 4139
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   4646 TTCACTGACGGAAGCAGCTATGTGGTGGAAGGTAAGAGGATGGCTGGGGCGGCGGTGGTG 4705

Qy   4140 gacgggacccgcacgatctgggccagcagcctgccggaaggaacttcagcgcaaaaggct 4199
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4706 GACGGGACCCACACGATCTGGGCCAGCAGCCTGCCGGAAGGAACTTCAGCGCAAAAGGCT 4765

Qy   4200 gagctcatggccctcacgcaagctttgcggctggccgaagggaaatccataaacatttat 4259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4766 GAGCTCATGGCCCTCACGCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAAACATTTAT 4825

Qy   4260 acggacagcaggtatgcctttgcgactgcacacgtacacggggccatctataaacaaagg 4319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4826 ACGGACAGCAGGTATGCCTTTGCGACTGCACACGTACACGGGGCCATCTATAAACAAAGG 4885

Qy   4320 gggttgcttacctcagcagggagggaaataaagaacaaagaggaaattctaagcctatta 4379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4886 GGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAGGAAATTCTAAGCCTATTA 4945

Qy   4380 gaagccttacatttgccaaaaaggctagctattatacactgtcctggacatcagaaagcc 4439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4946 GAAGCCTTACATTTGCCAAAAAGGCTAGCTATTATACACTGTCCTGGACATCAGAAAGCC 5005

Qy   4440 aaagatctcatatctagagggaaccagatggctgaccgggttgccaagcaggcagcccag 4499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5006 AAAGATCTCATATCTAGAGGGAACCAGATGGCTGACCGGGTTGCCAAGCAGGCAGCCCAG 5065
```


Qy 4500 gctgttaaccttctgcctataatagaaacgcccaaagccccagaacccagacgacagtac 4559
Db 5066 GCTGTTAACCTTCTGCCTATAATAGAAACGCCCAAAGCCCCAGAACCCAGACGACAGTAC 5125
Qy 4560 accctagaagactggcaagagataaaaaagatagaccagttctctgagactccggagggg 4619
Db 5126 ACCCTAGAAGACTGGCAAGAGATAAAAAAGATAGACCAGTTCTCTGAGACTCCGGAGGGG 5185
Qy 4620 acctgctatacctcatatgggaaggaaatcctgccccacaaagaagggttagaatatgtc 4679
Db 5186 ACCTGCTATACCTCATATGGGAAGGAAATCCTGCCCCACAAAGAAGGGTTAGAATATGTC 5245
Qy 4680 caacagatacatcgtctaacccacctaggaactaaacacctgcagcagttggtcagaaca 4739
Db 5246 CAACAGATACATCGTCTCACCCACCTAGGAACTAAACACCTGCAGCAGTTGGTCAGAACA 5305
Qy 4740 tcccctatcatgttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtg 4799
Db 5306 TCCCCTTATCATGTTCTGAGGCTACCAGGAGTGGCTGACTCGGTGGTCAAACATTGTGTG 5365
Qy 4800 ccctgccagctggttaatgctaatccttccagaatacctccaggaaagagactaagggga 4859
Db 5366 CCCTGCCAGCTGGTTAATGCTAATCCTTCCAGAATACCTCCAGGAAAGAGACTAAGGGGA 5425
Qy 4860 agccacccaggcgctcactgggaagtggacttcactgaggtaaagccggctaaatacgga 4919
Db 5426 AGCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAATACGGA 5485
Qy 4920 aacaaatatctattggtttttgtagacacctttcaggatgggtagaggcttatcctact 4979
Db 5486 AACAAATATCTATTGGTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACT 5545
Qy 4980 aagaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttccaagattt 5039
Db 5546 AAGAAAGAGACTTCAACCGTGGTGGCTAAAAAAATACTGGAGGAAATTTTTCCGAGATTT 5605
Qy 5040 ggaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcag 5099
Db 5606 GGAATACCTAAGGTAATCGGGTCAGACAATGGTCCAGCTTTTGTTGCCCAGGTAAGTCAG 5665
Qy 5100 ggactggccaagatattggggattgattggaaactgcattgtgcatacagaccccaaagc 5159
Db 5666 GGACTGGCCAAGATATTGGGGATTGATTGGAAACTGCATTGTGCATACAGACCCCAAAGC 5725
Qy 5160 tcaggacaggtagagaggatgaatagaaccattaaagagacccttaccaaattgaccaca 5219
Db 5726 TCAGGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACCCTTACCAAATTGACCACA 5785
Qy 5220 gagactggcattaatgattggatggctctcctgccctttgtgctttttagggtgaggaac 5279
Db 5786 GAGACTGGCATTAATGATTGGATAGCTCTCCTGCCCTTTGTGCTTTTTAGGGTTAGGAAC 5845
Qy 5280 acccctggacagtttgggctgacccccctatgaattgctctacgggggaccccccccgttg 5339
Db 5846 ACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTGCTCTACGGGGGACCCCCCCCGTTG 5905
Qy 5340 gcagaaattgcctttgcacatagtgctgatgtgctgctttcccagcctttgttctctagg 5399
Db 5906 GTAGAAATTGCTTCTGTACATAGTGCTGATGTGCTGCTTTCCCAGCCTCTGTTCTCTAGG 5965
Qy 5400 ctcaaggcgctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactca 5459
Db 5966 CTCAAGGCGCTCGAGTGGGTGAGGCAACGAGCGTGGAAGCAGCTCCGGGAGGCCTACTCA 6025
Qy 5460 gg---aggagacttgcaagttccacatcgcttccaagttggagattcagtctatgttaga 5516
Db 6026 GGAGAAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTGGGAGATTCAGTCTATGTTAGA 6085
Qy 5517 cgccaccgtgcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgacc 5576
Db 6086 CGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGCCCTTATCTCGTACTTTTGACC 6145
Qy 5577 acaccaacggctgtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaag 5636
Db 6146 ACACCAACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGATCCATACATCCCACGTTAAG 6205
Qy 5637 ccggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagctt 5696
Db 6206 CTGGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGACTGAGAATCCCCTTAAGCTT 6265
Qy 5697 cgcctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgcct 5756
Db 6266 CGCCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCT 6325
Qy 5757 tatagacagctcgaaccccccatagacctttatcccttacctggctgattattgaccctga 5816
Db 6326 TATAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGA 6385
Qy 5817 tacgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctga 5876
Db 6386 TACGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGTGGCCTGA 6445
Qy 5877 actgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagt 5936
Db 6446 ACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGCACACCTCCCAACCTAGT 6505
Qy 5937 ccgtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttc 5996
Db 6506 CCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGAGAAATACTGTGGGGATTC 6565
Qy 5997 tggggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatg 6056
Db 6566 TGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATG 6625
Qy 6057 gccgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagta 6116
Db 6626 GCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTA 6685
Qy 6117 caaaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaa 6176
Db 6686 CAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTATCTAAA 6745
Qy 6177 gataagtttcactg--aaaggaaaacaggaaaatattcaaaagtggataaatggtatgag 6234
Db 6746 GATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTATGAG 6805
Qy 6235 ctggggaatagttttttat-tatatggcgggggagcagggtccactttaaccattcgcct 6293
Db 6806 CTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTCCACTTTAACCATTCGCCT 6865
Qy 6294 taggatagagacggggacagaaccccctgtggcaatgggacccgataaagtactggctga 6353
Db 6866 TAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACCCGATAAAGTACTGGCTGA 6925
Qy 6354 acaggggccccggcccctggagccaccgcataacttgccggtgccccaattaacctcgct 6413
Db 6926 ACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCT 6985
Qy 6414 gcggcctgacataacacagccgcctagcaacagtaccactggattgattcctaccaacac 6473
Db 6986 GCGGCCTGACATAACACAGTCGCCTAGCAACGGTACCACTGGATTGATTCCTACCAACAC 7045
Qy 6474 gcctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatcca 6533
Db 7046 GCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCA 7105
Qy 6534 gggagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttg 6593
Db 7106 GGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTTG 7165
Qy 6594 tctatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgac 6653
Db 7166 TCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAGGAAAATTCAATGTGAC 7225
Qy 6654 caaagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagt 6713

```
Db    7226 CAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGT 7285

Qy    6714 ttccgggaaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatag 6773
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7286 TTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAG 7345

Qy    6774 tactgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtg 6833
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7346 TACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTG 7405

Qy    6834 gtgggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaa 6893
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7406 GTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAA 7465

Qy    6894 agatttctgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggt 6953
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7466 AGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGT 7525

Qy    6954 ccttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccct 7013
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7526 CCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCT 7585

Qy    7014 agctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgat 7073
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7586 AGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTGCCCTGAT 7645

Qy    7074 cacaggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaaga 7133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7646 CACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGA 7705

Qy    7134 tctccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctga 7193
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    7706 TCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGA 7765

Qy    7194 agtggttctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatg 7253
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7766 AGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATG 7825

Qy    7254 tgcagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactc 7313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7826 TGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTC 7885

Qy    7314 catgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggg 7373
           ||||| |||||||||| ||| ||||||||||||||||||||||||||||||||||||||||
Db    7886 CATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGG 7945

Qy    7374 gtggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgac 7433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7946 GTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGAC 8005

Qy    7434 ggggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtt 7493
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8006 GGGACCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTT 8065

Qy    7494 tgttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagta 7553
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8066 TGTTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTA 8125

Qy    7554 ccaaggccttctgagccaaggagaaactgacctctagccttcccagttctaagattagaa 7613
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8126 CCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAA 8185

Qy    7614 ctattaacaagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccaga 7673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8186 CTATTAACAAGACAAGAAGTGGGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGA 8245

Qy    7674 acccaggaagttaataaaaagctctaaatgcccccgaattacagaccctgctggctgcca 7733
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    8246 ACCCAGGAAGTTAATAAAAAGCTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCA 8305

Qy    7734 gtaaataggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagta 7793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8306 GTAAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTA 8365

Qy    7794 agataacaggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcacc 7853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8366 AGATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACC 8425

Qy    7854 atagaagaattgattacacattgacagccctagtgacctatctcaactgcaatctgtcac 7913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8426 ATAGAAGAATTGATTACACATTGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCAC 8485

Qy    7914 tctgcccaggagcccacgcagatgcggacctccggagctattttaaaatgattggtccac 7973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8486 TCTGCCCAGGAGCCCACGCAGATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCAC 8545

Qy    7974 ggagcgcgggctctcgatattttaaaatgattggtcca 8011
           |||||||||||||||||||||||||||||||||||||||
Db    8546 GGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCA 8583
```

RESULT 4
PEN133817

```
LOCUS       PEN133817    8849 bp    DNA             VRL       07-JAN-2000
DEFINITION  Porcine endogenous retrovirus type C proviral gag, pol and env
            genesand LTR (class A, clone 42).
ACCESSION   AJ133817
VERSION     AJ133817.1  GI:6688947
KEYWORDS    env gene; env protein; gag gene; gag protein; pol gene; pol
            protein.
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 8849)
  AUTHORS   Czauderna,F., Fischer,N., Boller,K., Krach,U., Kurth,R. and
            Toenjes,R.R.
  TITLE     Molecular Characterization of Human-tropic and
            Replication-competent Porcine Endogenous Retroviruses
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 8849)
  AUTHORS   Toenjes,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-1999) Toenjes R.R., Medical Biotechnology,
            Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, Hessen, D-63225
            Langen, GERMANY
FEATURES             Location/Qualifiers
     source          1. .8849
                     /organism="porcine endogenous retrovirus"
                     /proviral
                     /strain="type C"
                     /db_xref="taxon:61673"
                     /clone="42"
                     /note="grown in PK15 (ATCC No. CCL-33) cell line"
     LTR             1. .668
     misc_feature    506
                     /note="cap site"
     primer_bind     671. .688
                     /note="gly4"
     gene            1115. .2686
                     /gene="gag"
     mat_peptide     1115. .2686
                     /gene="gag"
                     /function="retroviral capsid"
                     /product="Gag protein"
     CDS             join(1115. .2686,2690. .6271,6150. .8132)
                     /note="viral polypeptide"
                     /codon_start=1
                     /protein_id="CAB65340.1"
                     /db_xref="GI:6688948"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPGSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPRIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
```

```
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREEREERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKDGRRGSDPLPEPRVTLKVEGQPVEFLVDTGAEHSVLLQPLGKL
                     KEKKSWVMGATGQRQYPWTTRRTVDLGVGRVTHSFLVIPECPVPLLGRDLLTKMGAQI
                     SFEQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMG
                     LAKQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPL
                     LPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFF
                     CLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFKIQH
                     PQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYS
                     LRGGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKE
                     KGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPW
                     RRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQ
                     PPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETG
                     VRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAE
                     LMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSL
                     LEALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPIIETPKAPEPR
                     RQYTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEGLEYVQQIHRLTHLGTKHLQ
                     QLVRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRIPPGKRLRGSHPGAHWEVDFTE
                     VKPAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKILEEIFPRFGIPKVIGSDNG
                     PAFVAQVSQGLAKILGIDWKLHCAYRPQSSGQVERMNRTIKETLTKLTTETGINDWMA
                     LLPFVLFRVRNTPGQFGLTPYELLYGGPPPLVEIASVHSADVLLSQPLFSRLKALEWV
                     RQRAWKQLREAYSGEGDLQVPHRFQVGDSVYVRRHRAGNLETRWKGPYLVLLTTPTAV
                     KVERIPTWIHASHVKPAPPPDSGWKAEKTENPLKLRLHRVVPYSVNNSSSMHPTLSRR
                     HLPIRGGKPKRLKIPLSFASIAWFLTLSITPQVNGKRLVDSPNSHKPLSLTWLLTDSG
                     TGININSTQGEAPLGTWWPELYVCLRSVIPGLNDQATPPDVLRAYGFYVCPGPPNNEE
                     YCGNPQDFFCKQWSCITSNDGNWKWPVSQQDRVSYSFVNNPTSYNQFNYGHGRWKDWQ
                     QRVQKDVRNKQISCHSLDLDYLKISFTEKGKQENIQKWVNGISWGIVYYGGSGRKKGS
                     VLTIRLRIETQMEPPVAIGPNKGLAEQGPPIQEQRPSPNPSDYNTTSGSVPTEPNITI
                     KTGAKLFSLIQGAFQALNSTTPEATSSCWLCLASGPPYYEGMARGGKFNVTKEHRDQC
                     TWGSQNKLTLTEVSGKGTCIGMVPPSHQHLCNHTEAFNRTSESQYLVPGYDRWWACNT
                     GLTPCVSTLVFNQTKDFCVMVQIVPRVYYYPEKAVLDKYDYRYNRPKREPISLTLAVM
                     LGLGVAAGVGTGTAALITGPQQLEKGLSNLHRIVTEDLQALEKSVSNLEESLTSLSEV
                     VLQNRRGLDLLFLKEGGLCVALKEECCFYVDHSGAIRDSMSKLRERLERRRREREADQ
                     GWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLINRFVAFVRERVSAVQIMVLR
                     QQYQGLLSQGETDL"
     mat_peptide     2690. .6271
                     /gene="pol"
                     /function="retroviral protease and polymerase"
                     /product="Pol protein"
     gene            2690. .6271
                     /gene="pol"
     mat_peptide     6150. .8129
                     /gene="env"
                     /function="retroviral envelope"
                     /product="Env protein"
     gene            6150. .8129
                     /gene="env"
     LTR             8182. .8849
     polyA_signal    8742. .8747
BASE COUNT     2355 a   2191 c   2281 g   2022 t
ORIGIN
```

Query Match 56.5%; Score 4640; DB 14; Length 8849;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 5330; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

```
Qy      1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    528 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 587

Qy     61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgagggggat 120
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    588 AAGACCGCTTCTTGTGAGTGATTTGGGGTGTCGCCTCTTCCGAGCCCGGACGAGGGGGAT 647

Qy    121 tgttcttttactggccttttcatttggtgcgttggccgggaaatcctgcgaccacccctta 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    648 TGTTCTTTTACTGGCCTTTCATTTGGTGCGTTGGCCGGGAAATCCTGCGACCACCCCTTA 707

Qy    181 cacccgagaaccgacttggaggtaaagggatccccctttggaacgtgtgtgtgtgtcggcc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    708 CACCCGAGAACCGACTTGGAGGTAAAGGGATCCCCTTTGGAACGTGTGTGTGTGTCGGCC 767

Qy    241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    768 GGCGTCTCTGTTCTGAGTGTCTGTTTTCGGTGATGCGCGCTTTCGGTTTGCAGCTGTCCT 827

Qy    301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    828 CTCAGACCGTAAGGACTGGAGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAGGCTG 887

Qy    361 ccaccctgggggacgcccgggaggtgggggagagccagggacgcctggtggtctcctact 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    888 CCACCCTGGGGGACGCCCCGGGAGGTGGGGAGAGCCAGGGACGCCTGGTGGTCTCCTACT 947

Qy    421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    948 GTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCG 1007

Qy    481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1008 ACTCTTTTGCCTGCTTGTGGAAGACGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGTTT 1067

Qy    541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1068 CTGTTTCGTGTGTCTTTGTCTTGTGCGTCCTTGTCTACAGTTTTAATATGGGACAGACAG 1127

Qy    601 tgactacccccctttagtttgactctcgaccattggactgaagttagatccagggctcata 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1128 TGACTACCCCCCTTAGTTTGACTCTCGACCATTGGACTGAAGTTAGATCCAGGGCTCATA 1187

Qy    661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1188 ATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGCCTCTGAATGGCCAA 1247

Qy    721 cattcgatgttggatggccatcagaggggaccttttaattctgaaattatcctggctgtta 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1248 CATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCTGAAATTATCCTGGCTGTTA 1307

Qy    781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1308 AGGCAATCATTTTTCAGACTGGACCCGGCTCTCATCCTGATCAGGAGCCCTATATCCTTA 1367

Qy    841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1368 CGTGGCAAGATTTGGCAGAAGATCCTCCGCCATGGGTTAAACCATGGCTAAATAAACCAA 1427

Qy    901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1428 GAAAGCCAGGTCCCCGAATCCTGGCTCTTGGAGAGAAAAACAAACACTCGGCCGAAAAAG 1487

Qy    961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1488 TCGAGCCCTC-TCCTCGTATCTACCCCGAGATCGAGGAGCCGCCGACTTGGCCGGAACCC 1546

Qy   1021 caacctgttcccccaccccctatccagcacagggtgctgtgagggga-cctctgcccct 1079
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   1547 CAACCTGTTCCCCCACCCCCTTATCCAGCACAGGGTGCTGTGAGGGGACCCTCTGCCCCT 1606

Qy   1080 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1607 CCTGGAGCTCCGGTGGTGGAGGGACCTGCTGCCGGGACTCGGAGCCGGAGAGGCGCCACC 1666

Qy   1140 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1667 CCGGAGCGGACAGACGAGATCGCGATATTACCGCTGCGCACCTATGGCCCTCCCATGCCA 1726

Qy   1200 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1727 GGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGG 1786

Qy   1260 aaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtcc 1319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1787 AAAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGAGTCC 1846
```

```
Qy   1320 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1847 CTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTC 1906

Qy   1380 acaaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgac 1439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1907 ACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAGAAAAAATGTTCCTGGGGCCGAC 1966

Qy   1440 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1967 GGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGATTTCCCTTGACTCGCCCCGGT 2026

Qy   1500 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2027 TGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCTGGTG 2086

Qy   1560 gcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatg 1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2087 GCGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAGGTGATG 2146

Qy   1620 cagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcgg 1679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2147 CAGGGACCGAACGAACCTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGG 2206

Qy   1680 ttcacccctttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcatt 1739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2207 TTCACCCCTTTTGATCCTACCTCAGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTCATT 2266

Qy   1740 gggcagtcggctctggatatcaggaagaaacttcagagactggaaggggttacaggaggct 1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2267 GGGCAGTCGGCTCTGGATATCAGAAAGAAACTTCAGAGACTGGAAGGGTTACAGGAGGCT 2326

Qy   1800 gagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagag 1859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2327 GAGTTACGTGATCTAGTGAGAGAGGCAGAGAAGGTGTATTACAGAAGGGAGACAGAAGAG 2386

Qy   1860 gagaaggaacagagaaaagaaaaggagagagaagaaagggaggaaaagacgtgatagacgg 1919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2387 GAGAAGGAACAGAGAAAAGAAAAGGAGAGAGAAGAAAGGGAGGAAAGACGTGATAGACGG 2446

Qy   1920 caagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggag 1979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2447 CAAGAGAAGAATTTGACTAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAG 2506

Qy   1980 agagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaat 2039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2507 AGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAAT 2566

Qy   2040 aggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagg 2099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2567 AGGACCCCACTCGACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGGACACTGGGCAAGG 2626

Qy   2100 aactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagat 2159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2627 AACTGCCCCAAGAAGGGAAACAAAGGACCGAAGGTCCTAGCTCTAGAAGAAGATAAAGAT 2686

Qy   2160 taggggagacggggttcggaccccctccccgagcccagggtaactttgaaggtggagggg 2219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2687 TAGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGAGGGG 2746

Qy   2220 caaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccatta 2279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2747 CAACCAGTTGAGTTCCTGGTTGATACCGGAGCGGAGCATTCAGTGCTGCTACAACCATTA 2806

Qy   2280 ggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatcca 2339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2807 GGAAAACTAAAAGAAAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACGGCAGTATCCA 2866

Qy   2340 tggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtc 2399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2867 TGGACTACCCGAAGAACCGTTGACTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTC 2926

Qy   2400 atccctgagtgcccagtacccccttctaggtagagacttactgaccaagatgggagctcaa 2459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2927 ATCCCTGAGTGCCCAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAA 2986

Qy   2460 atttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttg 2519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2987 ATTTCTTTTGAACAAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCACTGTGTTG 3046

Qy   2520 accctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagat 2579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3047 ACCCTCCAATTAGATGATGAATATCGACTATATTCTCCCCAAGTAAAGCCTGATCAAGAT 3106

Qy   2580 atacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttg 2639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3107 ATACAGTCCTGGTTGGAGCAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGGTTTG 3166

Qy   2640 gcaaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtc 2699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3167 GCAAAGCAAGTTCCCCCACAGGTTATTCAACTGAAGGCCAGTGCTACACCAGTATCAGTC 3226

Qy   2700 agacagtacccccttgagtagagaggctcgagaaggaatttggccgcatgttcaaagatta 2759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3227 AGACAGTACCCCTTGAGTAGAGAGGCTCGAGAAGGAATTTGGCCGCATGTTCAAAGATTA 3286

Qy   2760 atccaacagggcatcctagttcctgtccaatccccttggaatactcccctgctaccggtt 2819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3287 ATCCAACAGGGCATCCTAGTTCCTGTCCAATCCCCTTGGAATACTCCCCTGCTACCGGTT 3346

Qy   2820 aggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaagg 2879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3347 AGGAAGCCTGGGACCAATGATTATCGACCAGTACAGGACTTGAGAGAGGTCAATAAAAGG 3406

Qy   2880 gtgcaggacatacacccaacggtcccgaaccccttataacctcttgagcgccctcccgcct 2939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3407 GTGCAGGACATACACCCAACGGTCCCGAACCCTTATAACCTCTTGAGCGCCCTCCCGCCT 3466

Qy   2940 gaacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacac 2999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3467 GAACGGAACTGGTACACAGTATTGGACTTAAAAGATGCCTTCTTCTGCCTGAGATTACAC 3526

Qy   3000 cccactagccaaccacttttgccttcgaatggagagatccaggtacgggaagaaccggg 3059
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3527 CCCACTAGCCAACCACTTTTTGCCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGG 3586

Qy   3060 cagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaa 3119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3587 CAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGACGAA 3646

Qy   3120 gccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccag 3179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3647 GCCCTACACAGGGACCTGGCCAACTTCAAGATCCAACACCCTCAGGTGACCCTCCTCCAG 3706

Qy   3180 tacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaag 3239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3707 TACGTGGATGACCTGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGTACGAAG 3766

Qy   3240 gcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagatt 3299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3767 GCACTACTGCTGGAATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATT 3826

Qy   3300 tgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacg 3359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3827 TGCAGGAGAGAGGTAACATACTTGGGGTACAGTTTGCGGGGCGGGCAGCGATGGCTGACG 3886

Qy   3360 gaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgaga 3419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3887 GAGGCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAGTGAGA 3946

Qy   3420 gagtttttggggacagctggattttgcagactgtggatcccgggggtttgcgaccttagca 3479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3947 GAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTTAGCA 4006

Qy   3480 gccccactctacccgctaaccaaagaaaaaggggggattctcctgggctcctgagcaccag 3539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   4007 GCCCCACTCTACCCGCTAACCAAAGAAAAAGGGGAATTCTCCTGGGCTCCTGAGCACCAG 4066

Qy   3540 aaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctccctgac 3599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4067 AAGGCATTTGATGCTATCAAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCCTGAC 4126

Qy   3600 gtaactaaaccctttaccctttatgtggatgagcgtaagggagtagcccgaggagttttta 3659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4127 GTAACTAAACCCTTTACCCTTTATGTGGATGAGCGTAAGGGAGTAGCCCGAGGAGTTTTA 4186

Qy   3660 acccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagcttgatcct 3719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4187 ACCCAAACCCTAGGACCATGGAGGAGACCTGTTGCCTACCTGTCAAAGAAGCTTGATCCT 4246

Qy   3720 gtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccatactggtcaag 3779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4247 GTAGCCAGTGGTTGGCCCGTATGTCTGAAGGCTATCGCAGCTGTGGCCATACTGGTCAAG 4306

Qy   3780 gacgctgacaaattgactttgggacagaatataactgtaatagcccccatgcattggag 3839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4307 GACGCTGACAAATTGACTTTGGGACAGAATATAACTGTAATAGCCCCCCATGCATTGGAG 4366

Qy   3840 aacatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacccactatcaa 3899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4367 AACATCGTTCGGCAGCCCCCAGACCGATGGATGACCAACGCCCGCATGACCCACTATCAA 4426

Qy   3900 agcctgcttctcacagagagggtcactttcgctccaccagccgctctcaaccctgccact 3959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4427 AGCCTGCTTCTCACAGAGAGGGTCACTTTCGCTCCACCAGCCGCTCTCAACCCTGCCACT 4486

Qy   3960 cttctgcctgaagagactgatgaaccagtgactcatgattgccatcaactattgattgag 4019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4487 CTTCTGCCTGAAGAGACTGATGAACCAGTGACTCATGATTGCCATCAACTATTGATTGAG 4546

Qy   4020 gagactggggtccgcaaggaccttacagacataccgctgactggagaagtgctaacctgg 4079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4547 GAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGGAGAAGTGCTAACCTGG 4606

Qy   4080 ttcactgacggaagcagctatgtggtggaaggtaagaggatggctggggcggcagtggtg 4139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4607 TTCACTGACGGAAGCAGCTATGTGGTGGAAGGTAAGAGGATGGCTGGGGCGGCAGTGGTG 4666

Qy   4140 gacgggacccgcacgatctgggccagcagcctgccggaaggaacttcagcgcaaaaggct 4199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4667 GACGGGACCCGCACGATCTGGGCCAGCAGCCTGCCGGAAGGAACTTCAGCGCAAAAGGCT 4726

Qy   4200 gagctcatggccctcacgcaagctttgcggctggccgaagggaaatccataaacatttat 4259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4727 GAGCTCATGGCCCTCACGCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAAACATTTAT 4786

Qy   4260 acggacagcaggtatgcctttgcgactgcacacgtacacggggccatctataaacaaagg 4319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4787 ACGGACAGCAGGTATGCCTTTGCGACTGCACACGTACACGGGGCCATCTATAAACAAAGG 4846

Qy   4320 gggttgcttacctcagcagggagggaaataaagaacaaagaggaaattctaagcctatta 4379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4847 GGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAGGAAATTCTAAGCCTATTA 4906

Qy   4380 gaagccttacatttgccaaaaaggctagctattatacactgtcctggacatcagaaagcc 4439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4907 GAAGCCTTACATTTGCCAAAAAGGCTAGCTATTATACACTGTCCTGGACATCAGAAAGCC 4966

Qy   4440 aaagatctcatatctagagggaaccagatggctgaccgggttgccaagcaggcagcccag 4499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4967 AAAGATCTCATATCTAGAGGGAACCAGATGGCTGACCGGGTTGCCAAGCAGGCAGCCCAG 5026

Qy   4500 gctgttaaccttctgcctataatagaaacgcccaaagccccagaacccagacgacagtac 4559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5027 GCTGTTAACCTTCTGCCTATAATAGAAACGCCCAAAGCCCCAGAACCCAGACGACAGTAC 5086

Qy   4560 accctagaagactggcaagagataaaaaagatagaccagttctctgagactccggagggg 4619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5087 ACCCTAGAAGACTGGCAAGAGATAAAAAAGATAGACCAGTTCTCTGAGACTCCGGAGGGG 5146

Qy   4620 acctgctatacctcatatgggaaggaaatcctgccccacaaagaagggttagaatatgtc 4679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5147 ACCTGCTATACCTCATATGGGAAGGAAATCCTGCCCCACAAAGAAGGGTTAGAATATGTC 5206

Qy   4680 caacagatacatcgtctaacccacctaggaactaaacacctgcagcagttggtcagaaca 4739
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   5207 CAACAGATACATCGTCTCACCCACCTAGGAACTAAACACCTGCAGCAGTTGGTCAGAACA 5266

Qy   4740 tcccctatcatgttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtg 4799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5267 TCCCCTTATCATGTTCTGAGGCTACCAGGAGTGGCTGACTCGGTGGTCAAACATTGTGTG 5326

Qy   4800 ccctgccagctggttaatgctaatccttccagaatacctccaggaaagagactaagggga 4859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5327 CCCTGCCAGCTGGTTAATGCTAATCCTTCCAGAATACCTCCAGGAAAGAGACTAAGGGGA 5386

Qy   4860 agccacccaggcgctcactgggaagtggacttcactgaggtaaagccggctaaatacgga 4919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5387 AGCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAATACGGA 5446

Qy   4920 aacaaatatctattggtttttgtagacacctttcaggatgggtagaggcttatcctact 4979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5447 AACAAATATCTATTGGTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACT 5506

Qy   4980 aagaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttccaagattt 5039
          |||||||||||||||||||||||||| |||||||||||||||||||| ||||||
Db   5507 AAGAAAGAGACTTCAACCGTGGTGGCTAAAAAAATACTGGAGGAAATTTTTCCGAGATTT 5566

Qy   5040 ggaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcag 5099
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   5567 GGAATACCTAAGGTAATCGGGTCAGACAATGGTCCAGCTTTCGTTGCCCAGGTAAGTCAG 5626

Qy   5100 ggactggccaagatattggggattgattggaaactgcattgtgcatacagaccccaaagc 5159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5627 GGACTGGCCAAGATATTGGGGATTGATTGGAAACTGCATTGTGCATACAGACCCCAAAGC 5686

Qy   5160 tcaggacaggtagagaggatgaatagaaccattaaagagacccttaccaaattgaccaca 5219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5687 TCAGGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACCCTTACCAAATTGACCACA 5746

Qy   5220 gagactggcattaatgattggatggctctcctgccctttgtgctttttagggtgaggaac 5279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5747 GAGACTGGCATTAATGATTGGATGGCTCTCCTGCCCTTTGTGCTTTTTAGGGTGAGGAAC 5806

Qy   5280 acccctggacagtttgggctgacccccctatgaattgctctacgggggacccccccgttg 5339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5807 ACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTGCTCTACGGGGGACCCCCCCCGTTG 5866

Qy   5340 g 5340
          |
Db   5867 G 5867
```

```
RESULT   5
PEN293657
LOCUS       PEN293657    8763 bp    DNA             VRL       30-MAY-2001
DEFINITION  Porcine endogenous type C retrovirus proviral gag gene, pol gene
            and env gene, class B, clone 213.
ACCESSION   AJ293657
VERSION     AJ293657.1  GI:14275841
KEYWORDS    env gene; envelope; gag gene; group specific antigen; pol gene;
            polymerase; polyprotein.
SOURCE      porcine endogenous type C retrovirus.
  ORGANISM  porcine endogenous type C retrovirus
            Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
REFERENCE   1  (bases 1 to 8763)
  AUTHORS   Krach,U., Fischer,N., Czauderna,F. and Tonjes,R.R.
  TITLE     Comparison of replication-competent molecular clones of porcine
            endogenous retrovirus class a and class b derived from pig and
            human cells
  JOURNAL   J. Virol. 75 (12), 5465-5472 (2001)
  MEDLINE   21256017
```

```
REFERENCE   2  (bases 1 to 8763)
  AUTHORS   Toenjes,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2000) Toenjes R.R., Medical Biotechnology,
            Paul-Ehrlich-Institut, Paul-Ehrlich-Str. 51-95, Langen, 63225,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .8763
                     /organism="porcine endogenous type C retrovirus"
                     /proviral
                     /specific_host="Sus scrofa"
                     /db_xref="taxon:70540"
                     /clone="213"
                     /cell_line="PK15"
                     /note="class B"
     LTR             1. .630
     misc_feature    467
                     /note="Cap site"
     primer_bind     633. .650
     gene            1077. .2651
                     /gene="gag"
     CDS             1077. .2651
                     /gene="gag"
                     /codon_start=1
                     /product="group specific antigen"
                     /protein_id="CAC39625.1"
                     /db_xref="GI:14275842"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPGSHPDQEPYILTWQDLAEEPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPGIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREEEREERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKD"
     gene            2652. .6239
                     /gene="pol"
     CDS             <2652. .6239
                     /gene="pol"
                     /note="complete pol ORF, gag-pol is expressed as a
                     polyprotein"
                     /citation=[1]
                     /codon_start=1
                     /product="polymerase"
                     /protein_id="CAC39626.1"
                     /db_xref="GI:14275843"
                     /translation="GRRGSDPLPEPRVTLKVEGQPVEFLVDTGAEHSVLLQPLGKLKE
                     KKSWVMGATGQRQYPWTTRRTVDLAVGRVTHSFLVIPECPVPLLGRDLLTKMGAQISF
                     EQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLA
                     KQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLP
                     VRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCL
                     RLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTFFDEALHRDLANFRIQHPQ
                     VTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLR
                     GGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKG
                     EFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRR
                     PVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPP
                     DRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVR
                     KDLTDIPLTGEVLTWFTDGSSYVVEGKKMAGAAVVDGTRTIWASSLPEGTSAQKAELM
                     ALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLE
                     ALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPIIETPKAPEPRRQ
                     YTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEGLEYVQQIHRLTHLGTKHLQQL
                     VRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRIPPGKRLRGSHPGAHWEVDFTEVK
                     PAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKILEEIFPRFGIPKVIGSDNGPA
                     FVAQVSQGLAKILGIDWKLHCAYRPQSSGQVERMNRTIKETLTKLTTETGINDWIALL
                     PFVLFRVRNTPGQFGLTPYELLYGGPPPLVEIASVHSADVLLSQPLFSRLKALEWVRQ
                     RAWKQLREAYSGEGDLQVPHRFQVGDSVYVRRHHAGNLETRWKGPYLVLLTTPTAVKV
                     EGISTWIHASHVKLAPPPDSGWRAEKTENPLKLRLHRLVPYSNNNSPGQ"
     gene            6112. .8085
                     /gene="env"
     CDS             6112. .8085
                     /gene="env"
                     /codon_start=1
                     /product="envelope"
                     /protein_id="CAC39627.1"
                     /db_xref="GI:14275844"
                     /translation="MHPTLSWRHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSK
                     RLIDSSNPHRPLSLTWLIIDPDTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTP
                     PNLVRSYGFYCCPGTEKEKYCGGSGESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVN
                     SGPGKYKVMKLYKDKSCSPSDLDYLKISFTEKGKQENIQKWINGMSWGIVFYKYGGGA
                     GSTLTIRLRIETGTEPPVAVGPDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSX
                     STTGLIPTNTPRNSPGVPVKTGQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYY
                     EGMAKEGKFNVTKEHRNQCTWGSRNKLTLTEVSGKGTCIGKAPPSHQHLCNSTVVYEQ
                     ASENQYLVPGYNRWWACNTGLTPCVSTSVFNQSKDFCVMVQIVPRVYYYHPEEVVLDEY
                     DYRYNRPKREPVSLTLAVMLGLGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLR
                     ALEESVSNLEESLTSLSEVVLQNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDS
                     MSKLRERLERRRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLIN
                     RFVAFVRERVSAVQIMVLRQQYQGLLSQGETDL"
     LTR             8133. .8763
     polyA_signal    8655. .8660
BASE COUNT     2301 a    2200 c    2287 g    1974 t       1 others
ORIGIN
```

```
  Query Match             55.3%;  Score 4540;  DB 14;  Length 8763;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 7420;  Conservative    0;  Mismatches   42;  Indels    8;  Gaps    5;

Qy     434 ccgagttctgttgttgaagcgaaagcttccccctccgcggccgtccgactcttttgcctg 493
           ||||||
```

```
Qy   1153 acgagatcgcgatattaccgctgcgcacctatggccctcccatgccaggggccaattgc 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1642 ACGAGATCGCGATATTACCGCTGCGCACCTATGGCCCTCCCATGCCAGGGGGCCAATTGC 1701

Qy   1213 agcccctccagtattggccctttttcttctgcagatctctataattggaaaactaaccatc 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1702 AGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGGAAAACTAACCATC 1761

Qy   1273 ccccttttctcggaggatccccaacgcctcacggggttggtggagtcccttatgttctctc 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1762 CCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGAGTCCCTTATGTTCTCTC 1821

Qy   1333 accagcctacttgggatgattgtcaacagctgctgcagacactcttcacaaccgaggagc 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1822 ACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTCACAACCGAGGAGC 1881

Qy   1393 gagagagaattctgttagaggctagaaaaaatgttcctggggccgacgggcgacccacgc 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1882 GAGAGAGAATTCTGTTAGAGGCTAGAAAAAATGTTCCTGGGGCCGACGGGCGACCCACGC 1941

Qy   1453 agttgcaaaatgagattgacatgggatttcccttgactcgccccggttgggactacaaca 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1942 AGTTGCAAAATGAGATTGACATGGGATTTCCCTTGACTCGCCCCGGTTGGGACTACAACA 2001

Qy   1513 cggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcgggtctccggg 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2002 CGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCTGGTGGCGGGTCTCCGGG 2061

Qy   1573 gcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcagggaccgaacg 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2062 GCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAGGTGATGCAGGGACCGAACG 2121

Qy   1633 aacctccctcggtatttcttgagaggctcatggaagccttcaggcggttcacccctttg 1692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2122 AACCTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGGTTCACCCCTTTG 2181

Qy   1693 atcctacctcagaggcccagaaagcctcagtggccctggccttcattgggcagtcggctc 1752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2182 ATCCTACCTCAGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTCATTGGGCAGTCGGCTC 2241

Qy   1753 tggatatcaggaagaaacttcagagactggaagggttacaggaggctgagttacgtgatc 1812
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2242 TGGATATCAGAAAGAAACTTCAGAGACTGGAAGGGTTACAGGAGGCTGAGTTACGTGATC 2301

Qy   1813 tagtgagagaggcagagaaggtgtattacagaagggagacagaagaggagaaggaacaga 1872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2302 TAGTGAGAGAGGCAGAGAAGGTGTATTACAGAAGGGAGACAGAAGAGGAGAAGGAACAGA 2361

Qy   1873 gaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacggcaagagaagaatt 1932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2362 GAAAAGAAAAGGAGAGAGAAGAAAGGGAGGAAAGACGTGATAGACGGCAAGAGAAGAATT 2421

Qy   1933 tgactaagatcttggccgcagtggttgaagggaagagcagcagggagagagagagagatt 1992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2422 TGACTAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAGAGAGAGAGAGATT 2481

Qy   1993 ttaggaaaattaggtcaggccctagacagtcagggaacctgggcaataggaccccactcg 2052
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2482 TTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAATAGGACCCCACTCG 2541

Qy   2053 acaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaaggaactgccccaaga 2112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2542 ACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGGACACTGGGCAAGGAACTGCCCCAAGA 2601

Qy   2113 agggaaacaaaggaccgaaggtcctagctctagaagaagataaagattaggggagacggg 2172
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2602 AGGGAAACAAAGGACCGAAGGTCCTAGCTCTAGAAGAAGATAAAGATTAGGGGAGACGGG 2661

Qy   2173 gttcggacccccctccccgagcccagggtaactttgaaggtggagggggcaaccagttgagt 2232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2662 GTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGAGGGGCAACCAGTTGAGT 2721

Qy   2233 tcctggttgataccggagcggagcattcagtgctgctacaaccattaggaaaactaaaag 2292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2722 TCCTGGTTGATACCGGAGCGGAGCATTCAGTGCTGCTACAACCATTAGGAAAACTAAAAG 2781

Qy   2293 aaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatccatggactacccgaa 2352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2782 AAAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACGGCAGTATCCATGGACTACCCGAA 2841

Qy   2353 gaaccgttgacttgggagtgggacgggtaacccactcgtttctggtcatccctgagtgcc 2412
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   2842 GAACCGTTGACTTGGCAGTGGGACGGGTAACCCACTCGTTTCTGGTCATCCCTGAGTGCC 2901

Qy   2413 cagtacccccttctaggtagagacttactgaccaagatgggagctcaaatttcttttgaac 2472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2902 CAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAAATTTCTTTTGAAC 2961

Qy   2473 aaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttgaccctccaattag 2532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2962 AAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCACTGTGTTGACCCTCCAATTAG 3021

Qy   2533 atgatgaatatcgactatattctccccaagtaaagcctgatcaagatatacagtcctggt 2592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3022 ATGATGAATATCGACTATATTCTCCCCAAGTAAAGCCTGATCAAGATATACAGTCCTGGT 3081

Qy   2593 tggagcagtttccccaagcctgggcagaaaccgcagggatgggtttggcaaagcaagttc 2652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3082 TGGAGCAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGGTTTGGCAAAGCAAGTTC 3141

Qy   2653 ccccacaggttattcaactgaaggccagtgctacaccagtatcagtcagacagtacccct 2712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3142 CCCCACAGGTTATTCAACTGAAGGCCAGTGCTACACCAGTATCAGTCAGACAGTACCCCT 3201

Qy   2713 tgagtagagaggctcgagaaggaatttggccgcatgttcaaagattaatccaacagggca 2772
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3202 TGAGTAGAGAGGCTCGAGAAGGAATTTGGCCGCATGTTCAAAGATTAATCCAACAGGGCA 3261

Qy   2773 tcctagttcctgtccaatccccttggaatactcccctgctaccggttaggaagcctggga 2832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3262 TCCTAGTTCCTGTCCAATCCCCTTGGAATACTCCCCTGCTACCGGTTAGGAAGCCTGGGA 3321

Qy   2833 ccaatgattatcgaccagtacaggacttgagagaggtcaataaaagggtgcaggacatac 2892
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3322 CCAATGATTATCGACCAGTACAGGACTTGAGAGAGGTCAATAAAAGGGTGCAGGACATAC 3381

Qy   2893 acccaacggtcccgaacccttataacctcttgagcgccctcccgcctgaacggaactggt 2952
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3382 ACCCAACGGTCCCGAACCCTTATAACCTCTTGAGCGCCCTCCCGCCTGAACGGAACTGGT 3441

Qy   2953 acacagtattggacttaaaagatgccttcttctgcctgagattacaccccactagccaac 3012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3442 ACACAGTATTGGACTTAAAAGATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAAC 3501

Qy   3013 cacttttgccttcgaatggagagatccaggtacgggaagaaccgggcagctcacctgga 3072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3502 CACTTTTGCCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGA 3561

Qy   3073 cccgactgccccaagggttcaagaactccccgaccatctttgacgaagccctacacaggg 3132
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   3562 CCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACATTCTTTGACGAAGCCCTACACAGGG 3621

Qy   3133 acctggccaacttcaggatccaacaccctcaggtgaccctcctccagtacgtggatgacc 3192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3622 ACCTGGCCAACTTCAGGATCCAACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACC 3681

Qy   3193 tgcttctggcgggagccaccaaacaggactgcttagaaggtacgaaggcactactgctgg 3252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3682 TGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGTACGAAGGCACTACTGCTGG 3741

Qy   3253 aattgtctgacctaggctacagagcctctgctaagaaggcccagatttgcaggagagagg 3312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3742 AATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGG 3801

Qy   3313 taacatacttggggtacagtttgcggggcgggcagcgatggctgacggaggcacggaaga 3372
```

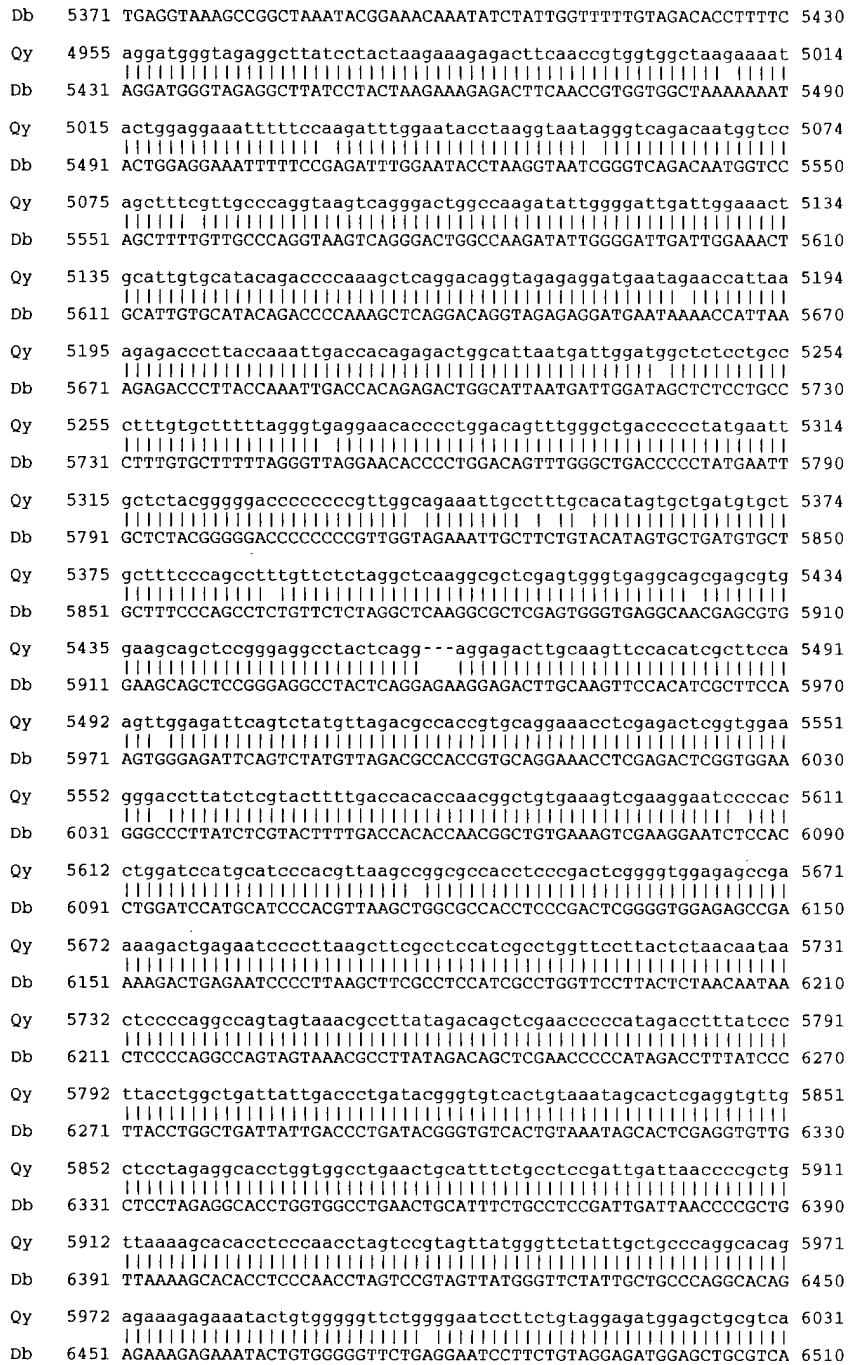
```
Db  3802 TAACATACTTGGGGTACAGTTTGCGGGGCGGGCAGCGATGGCTGACGGAGGCACGGAAGA 3861

Qy  3373 aaactgtagtccagataccggccccaaccacagccaaacaagtgagagagtttttgtgggga 3432

Db  3862 AAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAGTGAGAGAGTTTTTGGGGA 3921

Qy  3433 cagctggattttgcagactgtggatcccggggtttgcgaccttagcagccccactctacc 3492

Db  3922 CAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACC 3981

Qy  3493 cgctaaccaaagaaaaaggggattctcctgggctcctgagcaccagaaggcatttgatg 3552

Db  3982 CGCTAACCAAAGAAAAAGGGGAATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATG 4041

Qy  3553 ctatcaaaaaggccctgctgagcgcacctgctctggccctccctgacgtaactaaaccct 3612

Db  4042 CTATCAAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCT 4101

Qy  3613 ttacccttttatgtggatgagcgtaagggagtagcccgaggagttttaacccaaaccctag 3672

Db  4102 TTACCCTTTATGTGGATGAGCGTAAGGGAGTAGCCCGAGGAGTTTTAACCCAAACCCTAG 4161

Qy  3673 gaccatggaggagacctgttgcctacctgtcaaagaagcttgatcctgtagccagtggtt 3732

Db  4162 GACCATGGAGGAGACCTGTTGCCTACCTGTCAAAGAAGCTTGATCCTGTAGCCAGTGGTT 4221

Qy  3733 ggcccgtatgtctgaaggctatcgcagctgtggccatactggtcaaggacgctgacaaat 3792

Db  4222 GGCCCGTATGCCTGAAGGCTATCGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAAT 4281

Qy  3793 tgactttgggacagaatataactgtaatagccccccatgcattggagaacatcgttcggc 3852

Db  4282 TGACTTTGGGACAGAATATAACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGC 4341

Qy  3853 agcccccagaccgatggatgaccaacgcccgcatgacccactatcaaagcctgcttctca 3912

Db  4342 AGCCCCCAGACCGATGGATGACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCA 4401

Qy  3913 cagagagggtcactttcgctccaccagccgctctcaaccctgccactcttctgcctgaag 3972

Db  4402 CAGAGAGGGTCACTTTCGCTCCACCAGCCGCTCTCAACCCTGCCACTCTTCTGCCTGAAG 4461

Qy  3973 agactgatgaaccagtgactcatgattgccatcaactattgattgaggagactggggtcc 4032

Db  4462 AGACTGATGAACCAGTGACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCC 4521

Qy  4033 gcaaggaccttacagacataccgctgactggagaagtgctaacctggttcactgacggaa 4092

Db  4522 GCAAGGACCTTACAGACATACCGCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAA 4581

Qy  4093 gcagctatgtggtggaaggtaagaggatggctggggcggcagtggtggacgggacccgca 4152

Db  4582 GCAGCTATGTGGTGGAAGGTAAGAAGATGGCTGGAGCGGCGGTGGTGGACGGGACCCGCA 4641

Qy  4153 cgatctgggccagcagcctgccggaaggaacttcagcgcaaaaggctgagctcatggccc 4212

Db  4642 CGATCTGGGCCAGCAGCCTGCCGGAAGGAACTTCAGCGCAAAAGGCTGAGCTCATGGCCC 4701

Qy  4213 tcacgcaagctttgcggctggccgaagggaaatccataaacatttatacggacagcaggt 4272

Db  4702 TCACGCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAAACATTTATACGGACAGCAGGT 4761

Qy  4273 atgcctttgcgactgcacacgtacacggggccatctataaacaaaggggttgcttacct 4332

Db  4762 ATGCCTTTGCGACTGCACACGTACACGGGGCCATCTATAAACAAAGGGGGTTGCTTACCT 4821

Qy  4333 cagcagggagggaaataaagaacaaagaggaaattctaagcctattagaagccttacatt 4392

Db  4822 CAGCAGGGAGGGAAATAAAGAACAAAGAGGAAATTCTAAGCCTATTAGAAGCCTTACATT 4881

Qy  4393 tgccaaaaaggctagctattatacactgtcctggacatcagaaagccaaagatctcatat 4452

Db  4882 TGCCAAAAAGGCTAGCTATTATACACTGTCCTGGACATCAGAAAGCCAAAGATCTCATAT 4941

Qy  4453 ctagagggaaccagatggctgaccgggttgccaagcaggcagcccaggctgttaaccttc 4512

Db  4942 CTAGAGGGAACCAGATGGCTGACCGGGTTGCCAAGCAGGCAGCCCAGGCTGTTAACCTTC 5001

Qy  4513 tgcctataatagaaacgcccaaagccccagaacccagacgacagtacaccctagaagact 4572

Db  5002 TGCCTATAATAGAAACGCCCAAAGCCCCAGAACCCAGACGACAGTACACCCTAGAAGACT 5061

Qy  4573 ggcaagagataaaaaagatagaccagttctctgagactccggaggggacctgctatacct 4632

Db  5062 GGCAAGAGATAAAAAAGATAGACCAGTTCTCTGAGACTCCGGAGGGGACCTGCTATACCT 5121

Qy  4633 catatgggaaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatc 4692

Db  5122 CATATGGGAAGGAAATCCTGCCCCACAAAGAAGGGTTAGAATATGTCCAACAGATACATC 5181

Qy  4693 gtctaacccacctaggaactaaacacctgcagcagttggtcagaacatcccctatcatg 4752

Db  5182 GTCTCACCCACCTAGGAACTAAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATG 5241

Qy  4753 ttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctgg 4812

Db  5242 TTCTGAGGCTACCAGGAGTGGCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGG 5301

Qy  4813 ttaatgctaatccttccagaatacctccaggaaagagactaaggggaagccacccaggcg 4872

Db  5302 TTAATGCTAATCCTTCCAGAATACCTCCAGGAAAGAGACTAAGGGGAAGCCACCCAGGCG 5361

Qy  4873 ctcactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatctat 4932

Db  5362 CTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAATACGGAAACAAATATCTAT 5421

Qy  4933 tggtttttgtagacacctttcaggatgggtagaggcttatcctactaagaaagagactt 4992

Db  5422 TGGTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTT 5481

Qy  4993 caaccgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaagg 5052

Db  5482 CAACCGTGGTGGCTAAAAAAATACTGGAGGAAATTTTTCCGAGATTTGGAATACCTAAGG 5541

Qy  5053 taataggggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaaga 5112

Db  5542 TAATCGGGTCAGACAATGGTCCAGCTTTTGTTGCCCAGGTAAGTCAGGGACTGGCCAAGA 5601

Qy  5113 tattggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggtag 5172

Db  5602 TATTGGGGATTGATTGGAAACTGCATTGTGCATACAGACCCCAAAGCTCAGGACAGGTAG 5661

Qy  5173 agaggatgaatagaaccattaaagagaccccttaccaaattgaccacagagactggcatta 5232

Db  5662 AGAGGATGAATAGAACCATTAAAGAGACCCTTACCAAATTGACCACAGAGACTGGCATTA 5721

Qy  5233 atgattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacagt 5292

Db  5722 ATGATTGGATAGCTCTCCTGCCCTTTGTGCTTTTTAGGGTTAGGAACACCCCTGGACAGT 5781

Qy  5293 ttgggctgacccccctatgaattgctctacgggggacccccccccgttggcagaaattgcct 5352

Db  5782 TTGGGCTGACCCCCTATGAATTGCTCTACGGGGGACCCCCCCCGTTGGTAGAAATTGCTT 5841

Qy  5353 ttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcg 5412

Db  5842 CTGTACATAGTGCTGATGTGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGCGCTCG 5901

Qy  5413 agtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcagg---aggagact 5469

Db  5902 AGTGGGTGAGGCAACGAGCGTGGAAGCAGCTCCGGGAGGCCTACTCAGGAGAAGGAGACT 5961

Qy  5470 tgcaagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgcag 5529

Db  5962 TGCAAGTTCCACATCGCTTCCAAGTGGGAGATTCAGTCTATGTTAGACGCCACCATGCAG 6021
```

```
Qy   5530 gaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacggctg 5589
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   6022 GAAACCTCGAGACTCGGTGGAAGGGCCCTTATCTCGTACTTTTGACCACACCAACGGCTG 6081

Qy   5590 tgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccggcgccacctc 5649
          |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Db   6082 TGAAAGTCGAAGGAATCTCCACCTGGATCCATGCATCCCACGTTAAGCTGGCGCCACCTC 6141

Qy   5650 ccgactcggggtggagagccgaaaagactgagaatcccccttaagcttcgcctccatcgcc 5709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6142 CCGACTCGGGGTGGAGAGCCGAAAAGACTGAGAATCCCCTTAAGCTTCGCCTCCATCGCC 6201

Qy   5710 tggttccttactctaacaataactccccaggccagtagtaaacgccttatagacagctcg 5769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6202 TGGTTCCTTACTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTATAGACAGCTCG 6261

Qy   5770 aacccccatagacctttatcccttacctggctgattattgaccctgatacgggtgtcact 5829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6262 AACCCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGATACGGGTGTCACT 6321

Qy   5830 gtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaactgcatttctgc 5889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6322 GTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGTGGCCTGAACTGCATTTCTGC 6381

Qy   5890 ctccgattgattaaccccgctgttaaaagcacacctcccaacctagtccgtagttatggg 5949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6382 CTCCGATTGATTAACCCCGCTGTTAAAAGCACACCTCCCAACCTAGTCCGTAGTTATGGG 6441

Qy   5950 ttctattgctgcccaggcacagagaaagagaaatactgtgggggttctggggaatccttc 6009
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6442 TTCTATTGCTGCCCAGGCACAGAGAAAGAGAAATACTGTGGGGGTTCTGGGGAATCCTTC 6501

Qy   6010 tgtaggagatggagctgcgtcacctccaacgatggagactggaaatggccgatctctctc 6069
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6502 TGTAGGAGATGGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGCCGATCTCTCTC 6561

Qy   6070 caggaccgggtaaaattctcctttgtcaattccggcccgggcaagtacaaaatgatgaaa 6129
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   6562 CAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACAAAGTGATGAAA 6621

Qy   6130 ctatataaagataagagctgctccccatcagacttagattatctaaagataagtttcact 6189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6622 CTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTATCTAAAGATAAGTTTCACT 6681

Qy   6190 g--aaaggaaaacaggaaaatattcaaaagtggataaatggtatgagctggggaatagtt 6247
          |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6682 GAAAAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTATGAGCTGGGGAATAGTT 6741

Qy   6248 ttttat-tatatggcgggggagcagggtccactttaaccattcgccttaggatagagacg 6306
          ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6742 TTTTATAAATATGGCGGGGGAGCAGGGTCCACTTTAACCATTCGCCTTAGGATAGAGACG 6801

Qy   6307 gggacagaaccccctgtggcaatgggacccgataaagtactggctgaacaggggcccccg 6366
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   6802 GGGACAGAACCCCCTGTGGCAGTGGGACCCGATAAAGTACTGGCTGAACAGGGGCCCCCG 6861

Qy   6367 gccctggagccaccgcataacttgccggtgccccaattaacctcgctgcggcctgacata 6426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6862 GCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGCGGCCTGACATA 6921

Qy   6427 acacagccgcctagcaacagtaccactggattgattcctaccaacacgcctagaaactcc 6486
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   6922 ACACAGCCGCCTAGCANCAGTACCACTGGATTGATTCCTACCAACACGCCTAGAAACTCC 6981

Qy   6487 ccaggtgttcctgttaagacaggacagagactcttcagtctcatccagggagctttccaa 6546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6982 CCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGGGAGCTTTCCAA 7041

Qy   6547 gccatcaactccaccgaccctgatgccacttcttcttgttggctttgtctatcctcaggg 6606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7042 GCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGG 7101

Qy   6607 cctccttattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagcataga 6666
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   7102 CCTCCTTATTATGAGGGGATGGCTAAAGAAGGAAAATTCAATGTGACCAAAGAGCATAGA 7161

Qy   6667 aatcaatgtacatgggggtcccgaaataagcttaccctcactgaagtttccgggaagggg 6726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7162 AATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTTCCGGGAAGGGG 7221

Qy   6727 acatgcataggaaaagctcccccatcccaccaacaccctttgctatagtactgtggtttat 6786
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   7222 ACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCAATAGTACTGTGGTTTAT 7281

Qy   6787 gagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggcatgcaat 6846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7282 GAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGTGGGCATGCAAT 7341

Qy   6847 actgggttaaccccctgtgtttccacctcagtcttcaaccaatccaaagatttctgtgtc 6906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7342 ACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAGATTTCTGTGTC 7401

Qy   6907 atggtccaaatcgtccccgagtgtactaccatcctgaggaagtggtccttgatgaatat 6966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7402 ATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCCTTGATGAATAT 7461

Qy   6967 gactatcggtataaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctc 7026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7462 GACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAGCTGTAATGCTC 7521

Qy   7027 ggattagggacggccgttggcgtaggaacagggacagctgccctgatcacaggaccacag 7086
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7522 GGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTGCCCTGATCACAGGACCACAG 7581

Qy   7087 cagctagagaaaggacttggtgagctacatgcggccatgacagaagatctccgagccttа 7146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7582 CAGCTAGAGAAAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGATCTCCGAGCCTTA 7641

Qy   7147 aaggagtctgttagcaacctagaagagtccctgacttctttgtctgaagtggttctacag 7206
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7642 GAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAGTGGTTCTACAG 7701

Qy   7207 aaccggaggggattagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaa 7266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7702 AACCGGAGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAA 7761

Qy   7267 gaagaatgttgcttctatgtagatcactcaggagccatcagagactccatgaacaagctt 7326
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   7762 GAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTT 7821

Qy   7327 agaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaagga 7386
          ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7822 AGAGAAAGGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGA 7881

Qy   7387 tggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggcccctagta 7446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7882 TGGTTCAACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTA 7941

Qy   7447 gtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgccttttgtt 7506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7942 GTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTT 8001

Qy   7507 agagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctg 7566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8002 AGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTG 8061

Qy   7567 agccaaggagaaactgacctctagccttcccagttctaagattagaactattaacaagac 7626
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8062 AGCCAAGGAGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAACAAGAC 8121

Qy   7627 aagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagtta 7686
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8122 AAGAAGTGGGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTA 8181

Qy   7687 ataaaaagctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtaga 7746
```

```
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    8182 ATAAAAAGCTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGA 8241

Qy    7747 aggtcacacttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaa 7806
          |||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||
Db    8242 AGGTCACACTTGCTATTGTTCCAAGGCCTGCTATCCTGGCCTAAGTAAGATAACAGGAAA 8301

Qy    7807 tgagttgactaatcgcttatctggattctgtaaaactgactggcaccatagaagaattga 7866
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8302 TGAGTTGACTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACCATAGAAGAATTGA 8361

Qy    7867 ttacacattgacagccctagtgacctatct 7896
          |||||||||||||||||||||||||||||||
Db    8362 TTACACATTGACAGCCCTAGTGACCTATCT 8391


RESULT   6
PEN133818
LOCUS       PEN133818    8750 bp    DNA             VRL       14-APR-2000
DEFINITION  Porcine endogenous retrovirus type C proviral gag, pol and env
            genes and LTR (class B, clone 43).
ACCESSION   AJ133818
VERSION     AJ133818.1  GI:6688949
KEYWORDS    env gene; env protein; gag gene; gag protein; pol gene; pol
            protein.
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 8750)
  AUTHORS   Czauderna,F., Fischer,N., Boller,K., Kurth,R. and Tonjes,R.R.
  TITLE     Establishment and characterization of molecular clones of porcine
            endogenous retroviruses replicating on human cells
  JOURNAL   J. Virol. 74 (9), 4028-4038 (2000)
  MEDLINE   20219394
REFERENCE   2  (bases 1 to 8750)
  AUTHORS   Toenjes,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-1999) Toenjes R.R., Medical Biotechnology,
            Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, Hessen, D-63225
            Langen, GERMANY
FEATURES             Location/Qualifiers
     source          1. .8750
                     /organism="porcine endogenous retrovirus"
                     /proviral
                     /strain="type C"
                     /db_xref="taxon:61673"
                     /clone="43"
     LTR             1. .629
     misc_feature    467
                     /note="cap site"
     primer_bind     632. .649
                     /note="gly4"
     gene            1064. .2635
                     /gene="gag"
     mat_peptide     1064. .2635
                     /gene="gag"
                     /function="retroviral capsid"
                     /product="Gag protein"
     CDS             join(1064. .2635,2639. .6223,6099. .8072)
                     /note="viral polyprotein"
                     /codon_start=1
                     /protein_id="CAB65341.1"
                     /db_xref="GI:6688950"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPSSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPRIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREEREERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKDGRRGSDPLPEPRVALKVEGQPVEFLVDTGAEHSVLLQPLGKL
                     KEKKSWVMGATGQRQYPWTTRRTVDLAVGRVTHSFLVIPECPVPLLGRDLLTKMGAQI
                     SFEQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMG
                     LAKQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPL
                     LPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFF
                     CLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQH
                     PQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYS
                     LRGGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKE
                     KGEFSWAPEHQKTFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPW
                     RRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQ
                     PPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETG
                     VRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAE
                     LMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSL
                     LEALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPIIETPKAPEPR
                     RQYTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEGLEYVQQIHRLTHLGTKHLQ
                     QLVRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRVPPGKRLRGSHPGAHWEVDFTE
                     VKPAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKILEEIFPRFGIPKVIGSDNG
                     PAFVAQVSQGLAKILGIDWKLHCAYRPQSSGQVERMNKTIKETLTKLTTETGINDWIA
                     LLPFVLFRVRNTPGQFGLTPYELLYGGPPPLVEIASVHSADVLLSQPLFSRLKALEWV
                     RQRAWKQLREAYSGEGDLQVPHRFQVGDSVYVRRHRAGNLETRWKGPYLVLLTTPTAV
                     KVEGISTWIHASHVKLAPPPDSGWRAEKTENPLKLRLHRLVPYSNNNSPGQMHPTLSW
                     RHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSKRLIDSSNPHRPLSLTWLIIDP
                     DTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTPPNLVRSYGFYCCPGTEKEKYC
                     GGSEESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVNSGPGKYKVMKLYKDKSCSPSD
                     LDYLKISFTEKGKQENIQKWINGMSWGIVFYKYGGGAGSTLTIRLRIETGTEPPVAVG
                     PDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSNGTTGLIPTNTPRNSPGVPVKT
                     GQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYYEGMAKEGKFNVTKEHRNQCTW
                     GSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQASENQYLVPGYNRWWACNTGL
                     TPCVSSSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEYDYRYNRPKREPVSLTLAVMLG
                     LGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLRALEESVSNLEESLTSLSEVVL
                     QNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDSMSKLRERLERRRREREADQGW
                     FKGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLINRFVAFVRERVSAVQIMVLRQQ
                     YQGLLSQGETDL"
     mat_peptide     2639. .6223
                     /gene="pol"
                     /function="retroviral protease and polymerase"
                     /product="Pol protein"
     gene            2639. .6223
                     /gene="pol"
     mat_peptide     6099. .8069
                     /gene="env"
                     /function="retroviral envelope"
                     /product="Env protein"
     gene            6099. .8069
                     /gene="env"
     LTR             8122. .8750
     polyA_signal    8643. .8648
BASE COUNT     2288 a    2214 c    2286 g    1962 t
ORIGIN


  Query Match           55.2%;  Score 4532;  DB 14;  Length 8750;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 7862;  Conservative    0;  Mismatches   51;  Indels    8;  Gaps    5;

Qy     276 cgcgctttcggtttgcagctgtcctctcagaccgtaaggactggaggactgtgatcagca 335
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     752 CGCGCTTTCGGTTTGCAGCTGTCCTCTCAGACCGTAAGGACTGGAGGACTGTGATCAGCA 811

Qy     336 gacgtgctaggaggatcacaggctgccaccctgggggacgccccgggaggtggggagagc 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     812 GACGTGCTAGGAGGATCACAGGCTGCCACCCTGGGGGACGCCCCGGGAGGTGGGGAGAGC 871

Qy     396 cagggacgcctggtggtctcctactgtcggtcagaggaccgagttctgttgttgaagcga 455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     872 CAGGGACGCCTGGTGGTCTCCTACTGTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGA 931

Qy     456 aagcttccccctccgcggccgtccgactcttttgcctgcttgtggaagacgcggacgggt 515
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     932 AAGCTTCCCCCTCCGCGGCCGTCCGACTCTTTTGCCTGCTTGTGGAAGACGCGGACGGGT 991

Qy     516 cgcgtgtgtctggatctgttggtttctgtctcgtgtgtctttgtcttgtgcgtccttgtc 575
           || ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
```

```
Db    992 CGTGTGTGTCTGGATCTGTTGGTTTCTGTTTCGTGTGTCTTTGTCTTGTGCGTCCTTGTC 1051

Qy    576 tacagttttaatatgggacagacagtgactaccccccttagtttgactctcgaccattgg 635
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1052 TACAGTTTTAATATGGGACAGACAGTGACGACCCCCCTTAGTTTGACTCTCGACCATTGG 1111

Qy    636 actgaagttagatccagggctcataatttgtcagttcaggttaagaagggaccttggcag 695
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1112 ACTGAAGTTAGATCCAGGGCTCATAATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAG 1171

Qy    696 actttctgtgcctctgaatggccaacattcgatgttggatggccatcagaggggaccttt 755
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1172 ACTTTCTGTGCCTCTGAATGGCCAACATTCGATGTTGGATGGCCATCAGAGGGGACCTTT 1231

Qy    756 aattctgaaattatcctggctgttaaggcaatcatttttcagactggacccggctctcat 815
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   1232 AATTCTGAAATTATCCTGGCTGTTAAGGCAATCATTTTTCAGACTGGACCCAGCTCTCAT 1291

Qy    816 cctgatcaggagccctatatccttacgtggcaagatttggcagaagatcctccgccatgg 875
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1292 CCTGATCAGGAGCCCTATATCCTTACGTGGCAAGACTTGGCAGAAGATCCTCCGCCATGG 1351

Qy    876 gttaaaccatggctaaataaaccaagaaagccaggtccccgaatcctggctcttggagag 935
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1352 GTTAAACCATGGCTAAATAAACCAAGAAAGCCAGGTCCCCGAATCCTGGCTCTTGGAGAG 1411

Qy    936 aaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctaccccgagatcga 995
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1412 AAAAACAAACACTCGGCCGAAAAAGTCGAGCCCTC-TCCTCGTATCTACCCCGAGATCGA 1470

Qy    996 ggagccgccgacttggccggaaccccaacctgttcccccaccccctttatccagcacaggg 1055
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1471 GGAGCCGCCGACTTGGCCGGAACCCCAACCTGTTCCCCCACCCCCTTATCCAGCACAGGG 1530

Qy   1056 tgctgtgagggga-cctctgcccctcctggagctccggtggtggagggacctgctgccgg 1114
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   1531 TGCTGTGAGGGGACCCTCTGCCCCTCCTGGAGCTCCGGTGGTGGAGGGACCTGCTGCCGG 1590

Qy   1115 gactcggagccggagaggcgccaccccggagcggacagacgagatcgcgatattaccgct 1174
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1591 GACTCGGAGCCGGAGAGGCGCCACCCCGGAGCGGACAGACGAGATCGCGATATTACCGCT 1650

Qy   1175 gcgcacctatggccctcccatgccaggggggccaattgcagcccctccagtattggccctt 1234
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1651 GCGCACCTATGGCCCTCCCATGCCGGGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTT 1710

Qy   1235 ttcttctgcagatctctataattggaaaactaaccatccccctttctcggaggatccccca 1294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1711 TTCTTCTGCAGATCTCTATAATTGGAAAACTAACCATCCCCCTTTCTCGGAGGATCCCCA 1770

Qy   1295 acgcctcacggggttggtggagtcccttatgttctctcaccagcctacttgggatgattg 1354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1771 ACGCCTCACGGGGTTGGTGGAGTCCCTTATGTTCTCTCACCAGCCTACTTGGGATGATTG 1830

Qy   1355 tcaacagctgctgcagacactcttcacaaccgaggagcgagagagaattctgttagaggc 1414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1831 TCAACAGCTGCTGCAGACACTCTTCACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGC 1890

Qy   1415 tagaaaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatgagattgacat 1474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1891 TAGAAAAAATGTTCCTGGGGCCGACGGGCGACCCACGCAGTTGCAAAATGAGATTGACAT 1950

Qy   1475 gggatttcccttgactcgccccggttgggactacaacacggctgaaggtagggagagctt 1534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1951 GGGATTTCCCTTGACTCGCCCCGGTTGGGACTACAACACGGCTGAAGGTAGGGAGAGCTT 2010

Qy   1535 gaaaatctatcgccaggctctggtggcgggtctccggggcgcctcaagacggcccactaa 1594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2011 GAAAATCTATCGCCAGGCTCTGGTGGCGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAA 2070

Qy   1595 tttggctaaggtaagagaggtgatgcagggaccgaacgaacctccctcggtatttcttga 1654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2071 TTTGGCTAAGGTAAGAGAGGTGATGCAGGGACCGAACGAACCTCCCTCGGTATTTCTTGA 2130

Qy   1655 gaggctcatggaagccttcaggcggttcacccctttgatcctacctcagaggcccagaa 1714
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   2131 GAGGCTCATGGAAGCCTTCAGGCGGTTCACCCCTTTTGATCCTACCTCGGAGGCCCAGAA 2190

Qy   1715 agcctcagtggccctggccttcattgggcagtcggctctggatatcaggaagaaacttca 1774
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   2191 AGCCTCAGTGGCCCTGGCCTTCATTGGGCAGTCGGCTCTGGATATCAGAAAGAAACTTCA 2250

Qy   1775 gagactggaagggttacaggaggctgagttacgtgatctagtgagagaggcagagaaggt 1834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2251 GAGACTGGAAGGGTTACAGGAGGCTGAGTTACGTGATCTAGTGAGAGAGGCAGAGAAGGT 2310

Qy   1835 gtattacagaagggagacagaagaggagaaggaacagagaaaagaaaaggagagagaaga 1894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2311 GTATTACAGAAGGGAGACAGAAGAGGAGAAGGAACAGAGAAAAGAAAAGGAGAGAGAAGA 2370

Qy   1895 aagggaggaaagacgtgatagacggcaagagaagaatttgactaagatcttggccgcagt 1954
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   2371 AAGGGAGGAAAGACGTGATAGACGGCAAGAGAAAAATTTGACTAAGATCTTGGCCGCAGT 2430

Qy   1955 ggttgaagggaagagcagcagggagagagagagagattttaggaaaattaggtcaggccc 2014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2431 GGTTGAAGGGAAGAGCAGCAGGGAGAGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCC 2490

Qy   2015 tagacagtcagggaacctgggcaataggaccccactcgacaaggaccagtgtgcgtattg 2074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2491 TAGACAGTCAGGGAACCTGGGCAATAGGACCCCACTCGACAAGGACCAGTGTGCGTATTG 2550

Qy   2075 taaagaaaaaggacactgggcaaggaactgccccaagaagggaaacaaaggaccgaaggt 2134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2551 TAAAGAAAAAGGACACTGGGCAAGGAACTGCCCCAAGAAGGGAAACAAAGGACCGAAGGT 2610

Qy   2135 cctagctctagaagaagataaagattaggggagacggggttcggacccccctccccgagcc 2194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2611 CCTAGCTCTAGAAGAAGATAAAGATTAGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCC 2670

Qy   2195 cagggtaactttgaaggtggaggggcaaccagttgagttcctggttgataccggagcgga 2254
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2671 CAGGGTAGCTTTGAAGGTGGAGGGGCAACCAGTTGAGTTCCTGGTTGATACCGGAGCGGA 2730

Qy   2255 gcattcagtgctgctacaaccattaggaaaactaaaagaaaaaaaatcctgggtgatggg 2314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2731 GCATTCAGTGCTGCTACAACCATTAGGAAAACTAAAAGAAAAAAAATCCTGGGTGATGGG 2790

Qy   2315 tgccacagggcaacggcagtatccatggactacccgaagaaccgttgacttgggagtggg 2374
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   2791 TGCCACAGGGCAACGGCAGTATCCATGGACTACCCGAAGAACCGTTGACTTGGCAGTGGG 2850

Qy   2375 acgggtaacccactcgtttctggtcatccctgagtgcccagtacccctctaggtagaga 2434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2851 ACGGGTAACCCACTCGTTTCTGGTCATCCCTGAGTGCCCAGTACCCCTTCTAGGTAGAGA 2910

Qy   2435 cttactgaccaagatgggagctcaaatttcttttgaacaaggaagaccagaagtgtctgt 2494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2911 CTTACTGACCAAGATGGGAGCTCAAATTTCTTTTGAACAAGGAAGACCAGAAGTGTCTGT 2970

Qy   2495 gaataacaaacccatcactgtgttgaccctccaattagatgatgaatatcgactatattc 2554
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2971 GAATAACAAACCCATCACTGTGTTGACCCTCCAATTAGATGATGAATATCGACTATATTC 3030

Qy   2555 tccccaagtaaagcctgatcaagatatacagtcctggttggagcagtttccccaagcctg 2614
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3031 TCCCCAAGTAAAGCCTGATCAAGATATACAGTCCTGGTTGGAGCAGTTTCCCCAAGCCTG 3090

Qy   2615 ggcagaaaccgcagggatgggtttggcaaagcaagttcccccacaggttattcaactgaa 2674
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3091 GGCAGAAACCGCAGGGATGGGTTTGGCAAAGCAAGTTCCCCCACAGGTTATTCAACTGAA 3150

Qy   2675 ggccagtgctacaccagtatcagtcagacagtacccccttgagtagagaggctcgagaagg 2734
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3151 GGCCAGTGCTACACCAGTATCAGTCAGACAGTACCCCTTGAGTAGAGAGGCTCGAGAAGG 3210
```

```
Qy   2735 aatttggccgcatgttcaaagattaatccaacagggcatcctagttcctgtccaatcccc 2794
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3211 AATTTGGCCGCATGTTCAAAGATTAATCCAACAGGGCATCCTAGTTCCTGTCCAATCCCC 3270

Qy   2795 ttggaatactcccctgctaccggttaggaagcctgggaccaatgattatcgaccagtaca 2854
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3271 TTGGAATACTCCCCTGCTACCGGTTAGGAAGCCTGGGACCAATGATTATCGACCAGTACA 3330

Qy   2855 ggacttgagagaggtcaataaaagggtgcaggacatacacccaacggtcccgaacccttа 2914
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3331 GGACTTGAGAGAGGTCAATAAAAGGGTGCAGGACATACACCCAACGGTCCCGAACCCTTA 3390

Qy   2915 taacctcttgagcgccctcccgcctgaacggaactggtacacagtattggacttaaaaga 2974
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3391 TAACCTCTTGAGCGCCCTCCCGCCTGAACGGAACTGGTACACAGTATTGGACTTAAAAGA 3450

Qy   2975 tgccttcttctgcctgagattacacccccactagccaaccacttttttgccttcgaatggag 3034
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   3451 TGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAACCGCTTTTTGCCTTCGAATGGAG 3510

Qy   3035 agatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaa 3094
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3511 AGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAA 3570

Qy   3095 gaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatcca 3154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3571 GAACTCCCCGACCATCTTTGACGAAGCCCTACACAGGGACCTGGCCAACTTCAGGATCCA 3630

Qy   3155 acaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaa 3214
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3631 ACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACCAA 3690

Qy   3215 acaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctacag 3274
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3691 ACAGGACTGCTTAGAAGGTACGAAGGCACTACTGCTGGAATTGTCTGACCTAGGCTACAG 3750

Qy   3275 agcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagttt 3334
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3751 AGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGGTAACATACTTGGGGTACAGTTT 3810

Qy   3335 gcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggc 3394
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3811 GCGGGGCGGGCAGCGATGGCTGACGGAGGCACGGAAGAAAACTGTAGTCCAGATACCGGC 3870

Qy   3395 cccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagactgtg 3454
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3871 CCCAACCACAGCCAAACAAGTGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTG 3930

Qy   3455 gatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggggg 3514
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3931 GATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGGGA 3990

Qy   3515 attctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgag 3574
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   3991 ATTCTCCTGGGCTCCTGAGCACCAGAAGACATTTGATGCTATCAAAAAGGCCCTGCTGAG 4050

Qy   3575 cgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcg 3634
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4051 CGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCTTTACCCTTTATGTGGATGAGCG 4110

Qy   3635 taagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgttgc 3694
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4111 TAAGGGAGTAGCCCGAGGAGTTTTAACCCAAACCCTAGGACCATGGAGGAGACCTGTTGC 4170

Qy   3695 ctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctat 3754
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   4171 CTACCTGTCAAAGAAGCTTGATCCTGTAGCCAGTGGTTGGCCCGTATGCCTGAAGGCTAT 4230

Qy   3755 cgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatataac 3814
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4231 CGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAGAATATAAC 4290

Qy   3815 tgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatggatgac 3874
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4291 TGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAGACCGATGGATGAC 4350

Qy   3875 caacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgctcc 3934
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4351 CAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCACAGAGAGGGTCACTTTCGCTCC 4410

Qy   3935 accagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactca 3994
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4411 ACCAGCCGCTCTCAACCCTGCCACTCTTCTGCCTGAAGAGACTGATGAACCAGTGACTCA 4470

Qy   3995 tgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacatacc 4054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4471 TGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACC 4530

Qy   4055 gctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggtaa 4114
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4531 GCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGGTAA 4590

Qy   4115 gaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctgcc 4174
          |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||
Db   4591 GAGGATGGCTGGGGCGGCGGTGGTGGACGGGACCCACACGATCTGGGCCAGCAGCCTGCC 4650

Qy   4175 ggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggc 4234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4651 GGAAGGAACTTCAGCGCAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCTGGC 4710

Qy   4235 cgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcacacgt 4294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4711 CGAAGGGAAATCCATAAACATTTATACGGACAGCAGGTATGCCTTTGCGACTGCACACGT 4770

Qy   4295 acacggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaagaa 4354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4771 ACACGGGGCCATCTATAAACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAA 4830

Qy   4355 caaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctattat 4414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4831 CAAAGAGGAAATTCTAAGCCTATTAGAAGCCTTACATTTGCCAAAAAGGCTAGCTATTAT 4890

Qy   4415 acactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggctga 4474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4891 ACACTGTCCTGGACATCAGAAAGCCAAAGATCTCATATCTAGAGGGAACCAGATGGCTGA 4950

Qy   4475 ccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaa 4534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4951 CCGGGTTGCCAAGCAGGCAGCCCAGGCTGTTAACCTTCTGCCTATAATAGAAACGCCCAA 5010

Qy   4535 agccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagataga 4594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5011 AGCCCCAGAACCCAGACGACAGTACACCCTAGAAGACTGGCAAGAGATAAAAAAGATAGA 5070

Qy   4595 ccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaatcctgcc 4654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5071 CCAGTTCTCTGAGACTCCGGAGGGGACCTGCTATACCTCATATGGGAAGGAAATCCTGCC 5130

Qy   4655 ccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaa 4714
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   5131 CCACAAAGAAGGGTTAGAATATGTCCAACAGATACATCGTCTCACCCACCTAGGAACTAA 5190

Qy   4715 acacctgcagcagttggtcagaacatcccccttatcatgttctgaggctaccaggagtggc 4774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5191 ACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGGCTACCAGGAGTGGC 5250

Qy   4775 tgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaat 4834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   5251 TGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGGTTAATGCTAATCCTTCCAGAGT 5310

Qy   4835 acctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcac 4894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5311 ACCTCCAGGAAAGAGACTAAGGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTCAC 5370

Qy   4895 tgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctttc 4954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```


Db 5371 TGAGGTAAAGCCGGCTAAATACGGAAACAAATATCTATTGGTTTTTGTAGACACCTTTTC 5430
Qy 4955 aggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaat 5014
Db 5431 AGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAAAAAAT 5490
Qy 5015 actggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtcc 5074
Db 5491 ACTGGAGGAAATTTTTCCGAGATTTGGAATACCTAAGGTAATCGGGTCAGACAATGGTCC 5550
Qy 5075 agctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaact 5134
Db 5551 AGCTTTTGTTGCCCAGGTAAGTCAGGGACTGGCCAAGATATTGGGGATTGATTGGAAACT 5610
Qy 5135 gcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaa 5194
Db 5611 GCATTGTGCATACAGACCCCAAAGCTCAGGACAGGTAGAGAGGATGAATAAAACCATTAA 5670
Qy 5195 agagacccttaccaaattgaccacagagactggcattaatgattggatggctctcctgcc 5254
Db 5671 AGAGACCCTTACCAAATTGACCACAGAGACTGGCATTAATGATTGGATAGCTCTCCTGCC 5730
Qy 5255 ctttgtgctttttagggtgaggaacacccctggacagtttgggctgacccccctatgaatt 5314
Db 5731 CTTTGTGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAATT 5790
Qy 5315 gctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctgatgtgct 5374
Db 5791 GCTCTACGGGGGACCCCCCCCGTTGGTAGAAATTGCTTCTGTACATAGTGCTGATGTGCT 5850
Qy 5375 gctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcgagcgtg 5434
Db 5851 GCTTTCCCAGCCTCTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAACGAGCGTG 5910
Qy 5435 gaagcagctccgggaggcctactcagg---aggagacttgcaagttccacatcgcttcca 5491
Db 5911 GAAGCAGCTCCGGGAGGCCTACTCAGGAGAAGGAGACTTGCAAGTTCCACATCGCTTCCA 5970
Qy 5492 agttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaa 5551
Db 5971 AGTGGGAGATTCAGTCTATGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAA 6030
Qy 5552 gggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatccccac 5611
Db 6031 GGGCCCTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCCAC 6090
Qy 5612 ctggatccatgcatcccacgttaagccggcgccacctcccgactcggggtggagagccga 5671
Db 6091 CTGGATCCATGCATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGAGAGCCGA 6150
Qy 5672 aaagactgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataa 5731
Db 6151 AAAGACTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTACTCTAACAATAA 6210
Qy 5732 ctccccaggccagtagtaaacgccttatagacagctcgaaccccccatagacctttatccc 5791
Db 6211 CTCCCCAGGCCAGTAGTAAACGCCTTATAGACAGCTCGAACCCCCATAGACCTTTATCCC 6270
Qy 5792 ttacctggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttg 5851
Db 6271 TTACCTGGCTGATTATTGACCCTGATACGGGTGTCACTGTAAATAGCACTCGAGGTGTTG 6330
Qy 5852 ctcctagaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctg 5911
Db 6331 CTCCTAGAGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTG 6390
Qy 5912 ttaaaagcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacag 5971
Db 6391 TTAAAAGCACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAG 6450
Qy 5972 agaaagagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtca 6031
Db 6451 AGAAAGAGAAATACTGTGGGGGTTCTGAGGAATCCTTCTGTAGGAGATGGAGCTGCGTCA 6510
Qy 6032 cctccaacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcct 6091
Db 6511 CCTCCAACGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCT 6570
Qy 6092 ttgtcaattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgct 6151
Db 6571 TTGTCAATTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCT 6630
Qy 6152 ccccatcagacttagattatctaaagataagtttcactg--aaaggaaaacaggaaaata 6209
Db 6631 CCCCATCAGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATA 6690
Qy 6210 ttcaaaagtggataaatggtatgagctggggaatagttttttat-tatatggcgggggag 6268
Db 6691 TTCAAAAGTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAG 6750
Qy 6269 cagggtccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaa 6328
Db 6751 CAGGGTCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAG 6810
Qy 6329 tgggacccgataaagtactggctgaacaggggccccggccctggagccaccgcataact 6388
Db 6811 TGGGACCCGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACT 6870
Qy 6389 tgccggtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagta 6448
Db 6871 TGCCGGTGCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTA 6930
Qy 6449 ccactggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacag 6508
Db 6931 CCACTGGATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAG 6990
Qy 6509 gacagagactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctg 6568
Db 6991 GACAGAGACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTG 7050
Qy 6569 atgccacttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatgg 6628
Db 7051 ATGCCACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGG 7110
Qy 6629 ctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtccc 6688
Db 7111 CTAAAGAAGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCC 7170
Qy 6689 gaaataagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccc 6748
Db 7171 GAAATAAGCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCC 7230
Qy 6749 catcccaccaacaccttgctatagtactgtggtttatgagcaggcctcagaaaatcagt 6808
Db 7231 CATCCCACCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGT 7290
Qy 6809 atttagtacctggttataacaggtggtgggcatgcaatactgggttaaccccctgtgttt 6868
Db 7291 ATTTAGTACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTT 7350
Qy 6869 ccacctcagtcttcaaccaatccaaagatttctgtgtcatggtccaaatcgtcccccgag 6928
Db 7351 CCTCCTCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAG 7410
Qy 6929 tgtactaccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaa 6988
Db 7411 TGTACTACCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAA 7470
Qy 6989 aaagagaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcg 7048
Db 7471 AAAGAGAACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCG 7530
Qy 7049 taggaacagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtg 7108
Db 7531 TAGGAACAGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTG 7590

```
Qy   7109 agctacatgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctag 7168
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   7591 AGCTACATGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAG 7650

Qy   7169 aagagtccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgc 7228
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7651 AAGAGTCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGC 7710

Qy   7229 tgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtag 7288
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7711 TGTTTCTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAG 7770

Qy   7289 atcactcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtc 7348
          |||||||||||||||||||||||||||||| |||||||||| ||| ||||||||||||||
Db   7771 ATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTC 7830

Qy   7349 gaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttgga 7408
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   7831 GAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTAAAGGATGGTTCAACAGGTCTCCTTGGA 7890

Qy   7409 tgaccaccctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacag 7468
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   7891 TGACCACCCTGCTTTCTGCTCTGACGGGACCCCTAGTAGTCCTGCTCCTGTTACTTACAG 7950

Qy   7469 ttgggccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtcc 7528
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7951 TTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCC 8010

Qy   7529 agatcatggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctct 7588
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8011 AGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCT 8070

Qy   7589 agccttcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggat 7648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8071 AGCCTTCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTGGGGAATGAAAGGAT 8130

Qy   7649 gaaaatgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgccccc 7708
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8131 GAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAGCTCTAAATGCCCCC 8190

Qy   7709 gaattacagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttcc 7768
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8191 GAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTATTGTTCC 8250

Qy   7769 agggcctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatct 7828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8251 AGGGCCTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGACTAATCGCTTATCT 8310

Qy   7829 ggattctgtaaaactgactggcaccatagaagaattgattacacattgacagccctagtg 7888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8311 GGATTCTGTAAAACTGACTGGCACCATAGAAGAATTGATTACACATTGACAGCCCTAGTG 8370

Qy   7889 acctatctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccgg 7948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8371 ACCTATCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGACCTCCGG 8430

Qy   7949 agctattttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattggt 8008
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8431 AGCTATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGT 8490

Qy   8009 ccatggagcgcgggctctcgatattttaaaatgattggtttgtgacgcacaggctttgtt 8068
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8491 CCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTTTGTGACGCACAGGCTTTGTT 8550

Qy   8069 gtgaaccccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcg 8128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8551 GTGAACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGTCCTCTACCCCTGCG 8610

Qy   8129 tggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatca 8188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8611 TGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATCA 8670

Qy   8189 a 8189
```

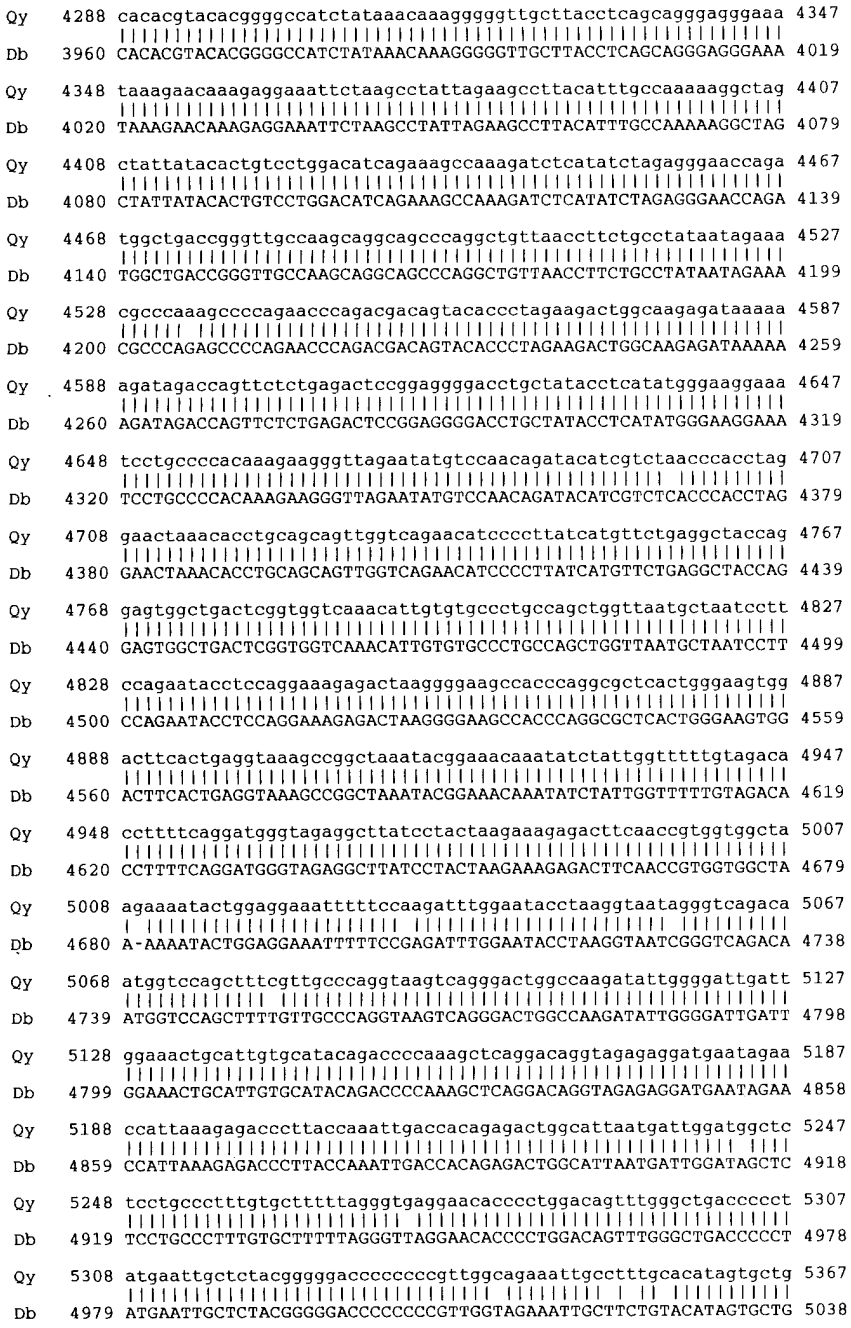
```
Db   8671 A 8671

RESULT   7
PERY17013
LOCUS       PERY17013     7808 bp    DNA             VRL       14-APR-2000
DEFINITION  Porcine endogenous retrovirus gag, pol, env genes.
ACCESSION   Y17013
VERSION     Y17013.1  GI:3005985
KEYWORDS    env gene; gag gene; pol gene.
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 7808)
  AUTHORS   Czauderna,F., Fischer,N., Boller,K., Kurth,R. and Tonjes,R.R.
  TITLE     Establishment and characterization of molecular clones of porcine
            endogenous retroviruses replicating on human cells
  JOURNAL   J. Virol. 74 (9), 4028-4038 (2000)
  MEDLINE   20219394
REFERENCE   2  (bases 1 to 7808)
  AUTHORS   Toenjes,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1998) R.R. Toenjes, Paul-Ehrlich Institut,
            Paul-Ehrlich Str. 51-59, P.O.Box 1740, D-63207 Langen, FRG
FEATURES             Location/Qualifiers
     source          1. .7808
                     /organism="porcine endogenous retrovirus"
                     /db_xref="taxon:61673"
                     /tissue_type="kidney"
                     /cell_line="PK15"
     gene            260. .1834
                     /gene="gag"
     CDS             260. .1834
                     /gene="gag"
                     /codon_start=1
                     /protein_id="CAA76581.1"
                     /db_xref="GI:3005986"
                     /db_xref="SPTREMBL:O73504"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPSSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPRIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTAEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRLGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKAYYGRETEEEKEQRKEKEREEREEERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKD"
     gene            2120. .4726
                     /gene="pol"
     CDS             2120. .4726
                     /gene="pol"
                     /codon_start=1
                     /protein_id="CAA76582.1"
                     /db_xref="GI:3005987"
                     /db_xref="SPTREMBL:O73505"
                     /translation="MGAQISFEQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQD
                     IQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQ
                     RLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLS
                     ALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNS
                     PTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYR
                     ASAKKAQICRREVTYLGYSLRGGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCR
                     LWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLY
                     VDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLT
                     LGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPE
                     ETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDG
                     THTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQR
                     GLLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQA
                     AQAVNLLPIIETPRAPEPRRQYTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEG
                     LEYVQQIHRLTHLGTKHLQQLVRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRIPP
                     GKRLRGSHPGAHWEVDFTEVKPAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKY
```

```
                     WRKFFRDLEYLR"
     gene            5294. .7267
                     /gene="env"
     CDS             5294. .7267
                     /gene="env"
                     /codon_start=1
                     /protein_id="CAA76583.1"
                     /db_xref="GI:3005988"
                     /db_xref="SPTREMBL:O73506"
                     /translation="MHPTLSWRHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSK
                     RLIDSSNPHRPLSLTWLIIDPDTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTP
                     PNLVRSYGFYCCPGTEKEKYCGGSEESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVN
                     SGPGKYKVMKLYKDKSCSPSDLDYLKISFTEKGKQENIQKWINGMSWGIVFYKYGGGA
                     GSTLTIRLRIETGTEPPVAVGPDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSN
                     GTTGLIPTNTPRNSPGVPVKTGQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYY
                     EGMAKEGKFNVTKEHRNQCTWGSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQ
                     ASENQYLVPGYNRWWACNTGLTPCVSSSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEY
                     DYRYNRPKREPVSLTLAVMLGLGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLR
                     ALEESVSNLEESLTSLSEVVLQNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDS
                     MSKLRERLERRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLIN
                     RFVAFVRERVGAVQIMVLRQQYQGLLSQGETDL"
BASE COUNT     2077 a    1971 c    2047 g    1713 t
ORIGIN
```

Query Match 49.9%; Score 4096; DB 14; Length 7808;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 7746; Conservative 0; Mismatches 55; Indels 9; Gaps 6;

```
Qy   329 atcagcagacgtgctaggaggatcacaggctgccaccctgggggacgccccgggaggtgg 388
Db     1 ATCAGCAGACGTGCTAGGAGGATCACAGGCTGCCACCCTGGGGGACGCCCCGGGAGGTGG 60

Qy   389 ggagagccagggacgcctggtggtctcctactgtcggtcagaggaccgagttctgttgtt 448
Db    61 GGAGAGCCAGGGACGCCTGGTGGTCTCCTACTGTCGGTCAGAGGACCGAGTTCTGTTGTT 120

Qy   449 gaagcgaaagcttcccccctccgcggccgtccgactcttttgcctgcttgtggaagacgcg 508
Db   121 GAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCGACTCTTTTGCCTGCTTGTGGAAGACGCG 180

Qy   509 gacgggtcgcgtgtgtctggatctgttggtttctgtctcgtgtgtctttgtcttgtgcgt 568
Db   181 GACGGGTCGTGTGTGTCTGGATCTGTTGGTTTCTGTTTCGTGTGTCTTTGTCTTGTGCGT 240

Qy   569 ccttgtctacagttttaatatgggacagacagtgactacccccttagtttgactctcga 628
Db   241 CCTTGTCTACAGTTTTAATATGGGACAGACAGTGACGACCCCCTTAGTTTAACTCTCGA 300

Qy   629 ccattggactgaagttagatccagggctcataatttgtcagttcaggttaagaagggacc 688
Db   301 CCATTGGACTGAAGTTAGATCCAGGGCTCATAATTTGTCAGTTCAGGTTAAGAAGGGACC 360

Qy   689 ttggcagactttctgtgcctctgaatggccaacattcgatgttggatggccatcagaggg 748
Db   361 TTGGCAGACTTTCTGTGCCTCTGAATGGCCAACATTCGATGTTGGATGGCCATCAGAGGG 420

Qy   749 gacctttaattctgaaattatcctggctgttaaggcaatcatttttcagactggacccgg 808
Db   421 GACCTTTAATTCTGAAATTATCCTGGCTGTTAAGGCAATCATTTTTCAGACTGGACCCAG 480

Qy   809 ctctcatcctgatcaggagccctatatccttacgtggcaagatttggcagaagatcctcc 868
Db   481 CTCTCATCCTGATCAGGAGCCCTATATCCTTACGTGGCAAGACTTGGCAGAAGATCCTCC 540

Qy   869 gccatgggttaaaccatggctaaataaaccaagaaagccaggtccccgaatcctggctct 928
Db   541 GCCATGGGTTAAACCATGGCTAAATAAACCAAGAAAGCCAGGTCCCCGAATCCTGGCTCT 600

Qy   929 tggagagaaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctaccccg 988
Db   601 TGGGGAGAAAAACAAACACTCGGCCGAAAAAGTCGAGCCCTC-TCCTCGTATCTACCCCG 659

Qy   989 agatcgaggagccgccgacttggccggaaccccaacctgttcccccacccccttatccag 1048
Db   660 AGATCGAGGAGCCGCCGACTTGGCCGGAACCCCAACCTGTTCCCCCACCCCCTTATCCAG 719

Qy  1049 cacagggtgctgtgagggga-cctctgcccctcctggagctccggtggtggagggacctg 1107
Db   720 CACAGGGTGCTGTGAGGGGACCCTCTGCCCCTCCTGGAGCTCCGGTGGTGGAGGGACCTG 779

Qy  1108 ctgccgggactcggagccggagaggcgccaccccggagcggacagacgagatcgcgatat 1167
Db   780 CTGCCGGGACTCGGAGCCGGAGAGGCGCCACCCCGGAGCGGACAGACGAGATCGCGATAT 839

Qy  1168 taccgctgcgcacctatggccctcccatgccaggggggccaattgcagcccctccagtatt 1227
Db   840 TACCGCTGCGCACCTATGGCCCTCCCATGCCGGGGGGCCAATTGCAGCCCCTCCAGTATT 899

Qy  1228 ggccctttttcttctgcagatctctataattggaaaaactaaccatccccctttctcggagg 1287
Db   900 GGCCCTTTTCTTCTGCAGATCTCTATAATTGGAAAACTAACCATCCCCCTTTCTCGGAGG 959

Qy  1288 atccccaacgcctcacggggttggtggagtcccttatgttctctcaccagcctacttggg 1347
Db   960 ATCCCCAACGCCTCACGGGGTTGGTGGAGTCCCTTATGTTCTCTCACCAGCCTACTTGGG 1019

Qy  1348 atgattgtcaacagctgctgcagacactcttcacaaccgaggagcgagagagaattctgt 1407
Db  1020 ATGATTGTCAACAGCTGCTGCAGACACTCTTCACAGCCGAGGAGCGAGAGAGAATTCTGT 1079

Qy  1408 tagaggctagaaaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatgaga 1467
Db  1080 TAGAGGCTAGAAAAAATGTTCCTGGGGCCGACGGGCGACCCACGCAGTTGCAAAATGAGA 1139

Qy  1468 ttgacatgggatttcccttgactcgccccggttgggactacaacacggctgaaggtaggg 1527
Db  1140 TTGACATGGGATTTCCCTTGACTCGCCTCGGTTGGGACTACAACACGGCTGAAGGTAGGG 1199

Qy  1528 agagcttgaaaatctatcgccaggctctggtggcgggtctccggggcgcctcaagacggc 1587
Db  1200 AGAGCTTGAAAATCTATCGCCAGGCTCTGGTGGCGGGTCTCCGGGGCGCCTCAAGACGGC 1259

Qy  1588 ccactaatttggctaaggtaagagaggtgatgcagggaccgaacgaacctccctcggtat 1647
Db  1260 CCACTAATTTGGCTAAGGTAAGAGAGGTGATGCAGGGACCGAACGAACCTCCCTCGGTAT 1319

Qy  1648 ttcttgagaggctcatggaagccttcaggcggttcaccccttttgatcctacctcagagg 1707
Db  1320 TTCTTGAGAGGCTCATGGAAGCCTTCAGGCGGTTCACCCCTTTTGATCCTACCTCGGAGG 1379

Qy  1708 cccagaaagcctcagtggccctggccttcattgggcagtcggctctggatatcaggaaga 1767
Db  1380 CCCAGAAAGCCTCAGTGGCCCTGGCCTTCATTGGGCAGTCGGCTCTGGATATCAGAAAGA 1439

Qy  1768 aacttcagagactggaagggttacaggaggctgagttacgtgatctagtgagagaggcag 1827
Db  1440 AACTTCAGAGACTGGAAGGGTTACAGGAGGCTGAGTTACGTGATCTAGTGAGAGAGGCAG 1499

Qy  1828 agaaggtgtattacagaagggagacagaagaggagaaggaacagagaaaagaaaaggaga 1887
Db  1500 AGAAGGCGTATTACGGAAGGGAGACAGAAGAGGAGAAGGAACAGAGAAAAGAAAAGGAGA 1559

Qy  1888 gagaagaaagggaggaaagacgtgatagacggcaagagaagaagaatttgactaagatcttgg 1947
Db  1560 GAGAAGAAAGGGAGGAAAGACGTGATAGACGGCAAGAGAAGAATTTGACTAAGATCTTGG 1619

Qy  1948 ccgcagtggttgaagggaagagcagcagggagagagagagagattttaggaaaattaggt 2007
Db  1620 CCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAGAGAGAGAGAGATTTTAGGAAAATTAGGT 1679

Qy  2008 caggccctagacagtcagggaacctgggcaataggaccccactcgacaaggaccagtgtg 2067
Db  1680 CAGGCCCTAGACAGTCAGGGAACCTGGGCAATAGGACCCCACTCGACAAGGACCAGTGTG 1739

Qy  2068 cgtattgtaaagaaaaaggacactgggcaaggaactgccccaagaagggaaacaaaggac 2127
```

```
Db  1740 CGTATTGTAAAGAAAAAGGACACTGGGCAAGGAACTGCCCCAAGAAGGGAAACAAAGGAC 1799
Qy  2128 cgaaggtcctagctctagaagaagataaagattaggggagacggggttcggacccccctcc 2187
Db  1800 CGAAGGTCCTAGCTCTAGAAGAAGATAAAGATTAGGGGAGACGGGGTTCGGACCCCCTCC 1859
Qy  2188 ccgagcccagggtaactttgaaggtggaggggcaaccagttgagttcctggttgataccg 2247
Db  1860 CTGAGCCCAGGGTAACTTTGAAGGTGGAGGGGCAACCAGTTGAGTTCCTGGTTGATACCG 1919
Qy  2248 gagcggagcattcagtgctgctacaaccattaggaaaactaaaagaaaaaaaatcctggg 2307
Db  1920 GAGCGGAGCATTCAGTGCTGCTACAACCATTAGGAAAACTAAAAGAAAAAAAATCCTGGG 1979
Qy  2308 tgatgggtgccacagggcaacggcagtatccatggactacccgaagaaccgttgacttgg 2367
Db  1980 TGGTGGGTGCCACAGGGCAACGGCAGTATCCATGGACTACCCGAAGAACCGTTGACTTGG 2039
Qy  2368 gagtgggacgggtaacccactcgtttctggtcatccctgagtgcccagtacccccttctag 2427
Db  2040 CAGTGGGACGGGTAACCCACTCGTTTCTGGTCATCCCTGAGTGCCCAGTACCCCTTCTAG 2099
Qy  2428 gtagagacttactgaccaagatgggagctcaaatttcttttgaacaaggaagaccagaag 2487
Db  2100 GTAGAGACTTACTGACCAAGATGGGAGCTCAAATTTCTTTTGAACAAGGAAGACCAGAAG 2159
Qy  2488 tgtctgtgaataacaaacccatcactgtgttgaccctccaattagatgatgaatatcgac 2547
Db  2160 TGTCTGTGAATAACAAACCCATCACTGTGTTGACCCTCCAATTAGATGATGAATATCGAC 2219
Qy  2548 tatattctccccaagtaaagcctgatcaagatatacagtcctggttggagcagtttcccc 2607
Db  2220 TATATTCTCCCCAAGTAAAGCCTGATCAAGATATACAGTCCTGGTTGGAGCAGTTTCCCC 2279
Qy  2608 aagcctgggcagaaaccgcagggatgggtttggcaaagcaagttcccccacaggttattc 2667
Db  2280 AAGCCTGGGCAGAAACCGCAGGGATGGGTTTGGCAAAGCAAGTTCCCCCACAGGTTATTC 2339
Qy  2668 aactgaaggccagtgctacaccagtatcagtcagacagtacccccttgagtagagaggctc 2727
Db  2340 AACTGAAGGCCAGTGCTACACCAGTATCAGTCAGACAGTACCCCTTGAGTAGAGAGGCTC 2399
Qy  2728 gagaaggaatttggccgcatgttcaaagattaatccaacagggcatcctagttcctgtcc 2787
Db  2400 GAGAAGGAATTTGGCCGCATGTTCAAAGATTAATCCAACAGGGCATCCTAGTTCCTGTCC 2459
Qy  2788 aatccccttggaatactcccctgctaccggttaggaagcctgggaccaatgattatcgac 2847
Db  2460 AATCCCCTTGGAATACTCCCCTGCTACCGGTTAGGAAGCCTGGGACCAATGATTATCGAC 2519
Qy  2848 cagtacaggacttgagagaggtcaataaaagggtgcaggacatacacccaacggtcccga 2907
Db  2520 CAGTACAGGACTTGAGAGAGGTCAATAAAAGGGTGCAGGACATACACCCAACGGTCCCGA 2579
Qy  2908 acccttataacctcttgagcgccctcccgcctgaacggaactggtacacagtattggact 2967
Db  2580 ACCCTTATAACCTCTTGAGCGCCCTCCCGCCTGAACGGAACTGGTACACAGTATTGGACT 2639
Qy  2968 taaaagatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcg 3027
Db  2640 TAAAAGATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAACCACTTTTTGCCTTCG 2699
Qy  3028 aatggagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaag 3087
Db  2700 AATGGAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTGCCCCAAG 2759
Qy  3088 ggttcaagaactccccgaccatctttgacgaagccctacacagggacctggccaacttca 3147
Db  2760 GGTTCAAGAACTCCCCGACCATCTTTGACGAAGCCCTACACAGGGACCTGGCCAACTTCA 2819
Qy  3148 ggatccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggag 3207
Db  2820 GGATCCAACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAG 2879
Qy  3208 ccaccaaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctag 3267
Db  2880 CCACCAAACAGGACTGCTTAGAAGGTACGAAGGCACTACTGCTGGAATTGTCTGACCTAG 2939
Qy  3268 gctacagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggt 3327
Db  2940 GCTACAGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGGTAACATACTTGGGGT 2999
Qy  3328 acagtttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccaga 3387
Db  3000 ACAGTTTGCGGGGCGGGCAGCGATGGCTGACGGAGGCACGGAAGAAAACTGTAGTCCAGA 3059
Qy  3388 taccggccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgca 3447
Db  3060 TACCGGCCCCAACCACAGCCAAACAAGTGAGAGAGTTTTTGGGGACAGCTGGATTTTGCA 3119
Qy  3448 gactgtggatcccgggggtttgcgaccttagcagccccactctacccgctaaccaaagaaa 3507
Db  3120 GACTGTGGATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAA 3179
Qy  3508 aagggggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccc 3567
Db  3180 AAGGGGAATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATGCTATCAAAAAGGCCC 3239
Qy  3568 tgctgagcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtgg 3627
Db  3240 TGCTGAGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCTTTACCCTTTATGTGG 3299
Qy  3628 atgagcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagac 3687
Db  3300 ATGAGCGTAAGGGAGTAGCCCGAGGAGTTTTAACCCAAACCCTAGGACCATGGAGGAGAC 3359
Qy  3688 ctgttgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctga 3747
Db  3360 CTGTTGCCTACCTGTCAAAGAAGCTTGATCCTGTAGCCAGTGGTTGGCCCGTATGCCTGA 3419
Qy  3748 aggctatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacaga 3807
Db  3420 AGGCTATCGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAATTGACCTTGGGACAGA 3479
Qy  3808 atataactgtaatagccccccatgcattggagaacatcgttcggcagccccagaccgat 3867
Db  3480 ATATAACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAGACCGAT 3539
Qy  3868 ggatgaccaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactt 3927
Db  3540 GGATGACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCACAGAGAGGGTCACTT 3599
Qy  3928 tcgctccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccag 3987
Db  3600 TCGCTCCACCAGCCGCTCTCAACCCTGCCACTCTTCTGCCTGAAGAGACTGATGAACCAG 3659
Qy  3988 tgactcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttacag 4047
Db  3660 TGACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAG 3719
Qy  4048 acataccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtgg 4107
Db  3720 ACATACCGCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGG 3779
Qy  4108 aaggtaagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagca 4167
Db  3780 AAGGTAAGAGGATGGCTGGGGCGGCGGTGGTGGACGGGACCCACACGATCTGGGCCAGCA 3839
Qy  4168 gcctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgc 4227
Db  3840 GCCTGCCGGAAGGAACTTCAGCGCAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGC 3899
Qy  4228 ggctggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactg 4287
Db  3900 GGCTGGCCGAAGGGAAATCCATAAACATTTATACGGACAGCAGGTATGCCTTTGCGACTG 3959
```


Qy 4288 cacacgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaa 4347
Db 3960 CACACGTACACGGGGCCATCTATAAACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAA 4019
Qy 4348 taaagaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctag 4407
Db 4020 TAAAGAACAAAGAGGAAATTCTAAGCCTATTAGAAGCCTTACATTTGCCAAAAAGGCTAG 4079
Qy 4408 ctattatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccaga 4467
Db 4080 CTATTATACACTGTCCTGGACATCAGAAAGCCAAAGATCTCATATCTAGAGGGAACCAGA 4139
Qy 4468 tggctgaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaa 4527
Db 4140 TGGCTGACCGGGTTGCCAAGCAGGCAGCCCAGGCTGTTAACCTTCTGCCTATAATAGAAA 4199
Qy 4528 cgcccaaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaa 4587
Db 4200 CGCCCAGAGCCCCAGAACCCAGACGACAGTACACCCTAGAAGACTGGCAAGAGATAAAAA 4259
Qy 4588 agatagaccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaa 4647
Db 4260 AGATAGACCAGTTCTCTGAGACTCCGGAGGGGACCTGCTATACCTCATATGGGAAGGAAA 4319
Qy 4648 tcctgccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctag 4707
Db 4320 TCCTGCCCCACAAAGAAGGGTTAGAATATGTCCAACAGATACATCGTCTCACCCACCTAG 4379
Qy 4708 gaactaaacacctgcagcagttggtcagaacatcccccttatcatgttctgaggctaccag 4767
Db 4380 GAACTAAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGGCTACCAG 4439
Qy 4768 gagtggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatcctt 4827
Db 4440 GAGTGGCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGGTTAATGCTAATCCTT 4499
Qy 4828 ccagaatacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtgg 4887
Db 4500 CCAGAATACCTCCAGGAAAGAGACTAAGGGGAAGCCACCCAGGCGCTCACTGGGAAGTGG 4559
Qy 4888 acttcactgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagaca 4947
Db 4560 ACTTCACTGAGGTAAAGCCGGCTAAATACGGAAACAAATATCTATTGGTTTTTGTAGACA 4619
Qy 4948 ccttttcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggcta 5007
Db 4620 CCTTTTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTA 4679
Qy 5008 agaaaatactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagaca 5067
Db 4680 A-AAAATACTGGAGGAAATTTTTCCGAGATTTGGAATACCTAAGGTAATCGGGTCAGACA 4738
Qy 5068 atggtccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgatt 5127
Db 4739 ATGGTCCAGCTTTTGTTGCCCAGGTAAGTCAGGGACTGGCCAAGATATTGGGGATTGATT 4798
Qy 5128 ggaaactgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaa 5187
Db 4799 GGAAACTGCATTGTGCATACAGACCCCAAAGCTCAGGACAGGTAGAGAGGATGAATAGAA 4858
Qy 5188 ccattaaagagacccttaccaaattgaccacagagactggcattaatgattggatggctc 5247
Db 4859 CCATTAAAGAGACCCTTACCAAATTGACCACAGAGACTGGCATTAATGATTGGATAGCTC 4918
Qy 5248 tcctgccctttgtgcttttagggtgaggaacacccctggacagtttgggctgacccccct 5307
Db 4919 TCCTGCCCTTTGTGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCT 4978
Qy 5308 atgaattgctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctg 5367
Db 4979 ATGAATTGCTCTACGGGGGACCCCCCCCGTTGGTAGAAATTGCTTCTGTACATAGTGCTG 5038
Qy 5368 atgtgctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagc 5427
Db 5039 ATGTGCTGCTTTCCCAGCCTCTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAAC 5098
Qy 5428 gagcgtggaagcagctccgggaggcctactcagg---aggagacttgcaagttccacatc 5484
Db 5099 GAGCGTGGAAGCAGCTCCGGGAGGCCTACTCAGGAGAAGGAGACTTGCAAGTTCCACATC 5158
Qy 5485 gcttccaagttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactc 5544
Db 5159 GCTTCCAAGTGGGAGATTCAGTCTATGTTAGACGCCACCGTGCAGGAAACCTCGAGACTC 5218
Qy 5545 ggtggaagggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaa 5604
Db 5219 GGTGGAAGGGCCCTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAA 5278
Qy 5605 tccccacctggatccatgcatcccacgttaagccggcgccacctcccgactcggggtgga 5664
Db 5279 TCTCCACCTGGATCCATGCATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGA 5338
Qy 5665 gagccgaaaagactgagaatccccttaagcttcgcctccatcgcctggttccttactcta 5724
Db 5339 GAGCCGAAAAGACTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTACTCTA 5398
Qy 5725 acaataactccccaggccagtagtaaacgccttatagacagctcgaaccccccatagacct 5784
Db 5399 ACAATAACTCCCCAGGCCAGTAGTAAACGCCTTATAGACAGCTCGAACCCCCATAGACCT 5458
Qy 5785 ttatcccttacctggctgattattgaccctgatacgggtgtcactgtaaatagcactcga 5844
Db 5459 TTATCCCTTACCTGGCTGATTATTGACCCTGATACGGGTGTCACTGTAAATAGCACTCGA 5518
Qy 5845 ggtgttgctcctagaggcacctggtggcctgaactgcatttctgcctccgattgattaac 5904
Db 5519 GGTGTTGCTCCTAGAGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAAC 5578
Qy 5905 cccgctgttaaaagcacacctcccaacctagtccgtagttatgggttctattgctgccca 5964
Db 5579 CCCGCTGTTAAAAGCACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCA 5638
Qy 5965 ggcacagagaaagagaaatactgtgggggttctggggaatccttctgtaggagatggagc 6024
Db 5639 GGCACAGAGAAAGAGAAATACTGTGGGGGTTCTGAGGAATCCTTCTGTAGGAGATGGAGC 5698
Qy 6025 tgcgtcacctccaacgatggagactggaaatggccgatctctctccaggaccgggtaaaa 6084
Db 5699 TGCGTCACCTCCAACGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAA 5758
Qy 6085 ttctcctttgtcaattccggcccgggcaagtacaaaatgatgaaactatataaagataag 6144
Db 5759 TTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAG 5818
Qy 6145 agctgctccccatcagacttagattatctaaagataagtttcactg--aaaggaaaacag 6202
Db 5819 AGCTGCTCCCCATCAGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAG 5878
Qy 6203 gaaaatattcaaaagtggataaatggtatgagctggggaatagttttttat-tatatggc 6261
Db 5879 GAAAATATTCAAAAGTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGC 5938
Qy 6262 gggggagcagggtccactttaaccattcgccttaggatagagacggggacagaaccccct 6321
Db 5939 GGGGGAGCAGGGTCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCT 5998
Qy 6322 gtggcaatgggacccgataaagtactggctgaacaggggccccccggccctggagccaccg 6381
Db 5999 GTGGCAGTGGGACCCGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCG 6058
Qy 6382 cataacttgccggtgccccaattaacctcgctgcggcctgacataacacagccgcctagc 6441
Db 6059 CATAACTTGCCGGTGCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGC 6118
Qy 6442 aacagtaccactggattgattcctaccaacacgcctagaaactccccaggtgttcctgtt 6501

```
Db   6119 AACGGTACCACTGGATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTT 6178

Qy   6502 aagacaggacagagactcttcagtctcatccagggagctttccaagccatcaactccacc 6561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6179 AAGACAGGACAGAGACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACC 6238

Qy   6562 gaccctgatgccacttcttcttgttggctttgtctatcctcagggcctccttattatgag 6621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6239 GACCCTGATGCCACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAG 6298

Qy   6622 gggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgg 6681
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   6299 GGGATGGCTAAAGAAGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGG 6358

Qy   6682 gggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgcataggaaaa 6741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6359 GGGTCCCGAAATAAGCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAA 6418

Qy   6742 gctcccccatcccaccaacacctttgctatagtactgtggtttatgagcaggcctcagaa 6801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6419 GCTCCCCCATCCCACCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAA 6478

Qy   6802 aatcagtatttagtacctggttataacaggtggtgggcatgcaatactgggttaaccccc 6861
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6479 AATCAGTATTTAGTACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCC 6538

Qy   6862 tgtgtttccacctcagtcttcaaccaatccaaagatttctgtgtcatggtccaaatcgtc 6921
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6539 TGTGTTTCCTCCTCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTC 6598

Qy   6922 ccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcggtataac 6981
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6599 CCCCGAGTGTACTACCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAAC 6658

Qy   6982 cgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggattagggacggcc 7041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6659 CGACCAAAAAGAGAACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCC 6718

Qy   7042 gttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaaagga 7101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6719 GTTGGCGTAGGAACAGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGA 6778

Qy   7102 cttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgttagc 7161
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   6779 CTTGGTGAGCTACATGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGC 6838

Qy   7162 aacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggaggggatta 7221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6839 AACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTA 6898

Qy   7222 gatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttc 7281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6899 GATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTC 6958

Qy   7282 tatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagttagag 7341
          |||||||||||||||||||||||||||||||||||| |||||||||| ||| |||||||
Db   6959 TATGTAGATCACTCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAG 7018

Qy   7342 aggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtct 7401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7019 AGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCT 7078

Qy   7402 ccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctcctgtta 7461
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   7079 CCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGACCCCTAGTAGTCCTGCTCCTGTTA 7138

Qy   7462 cttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacgagtgagt 7521
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   7139 CTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAGTGGGT 7198

Qy   7522 gcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaaggagaaact 7581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   7199 GCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACT 7258

Qy   7582 gacctctagccttcccagttctaagattagaactattaacaagacaagaagtggggaatg 7641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7259 GACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTGGGGAATG 7318

Qy   7642 aaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaa 7701
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7319 AAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAGCTCTAAA 7378

Qy   7702 tgcccccgaattacagaccctgctggctgccagtaaataggtagaaggtcacacttccta 7761
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   7379 TGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTA 7438

Qy   7762 ttgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttgactaatcg 7821
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   7439 TTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGCTGACTAATCG 7498

Qy   7822 cttatctggattctgtaaaactgactggcaccatagaagaattgattacacattgacagc 7881
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7499 CTTATCTGGATTCTGTAAAACTGACTGGCACCATAGAAGAATTGATTACACATTGACAGC 7558

Qy   7882 cctagtgacctatctcaactgcaatctgtcactctgcccaggagcccacgcagatgcgga 7941
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7559 CCTAGTGACCTATCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGA 7618

Qy   7942 cctccggagctattttaaaatgattggtccacggagcgcgggctctcgatattttaaaat 8001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7619 CCTCCGGAGCTATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAAT 7678

Qy   8002 gattggtccatggagcgcgggctctcgatattttaaaatgattggtttgtgacgcacagg 8061
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7679 GATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTTTGTGACGCACAGG 7738

Qy   8062 ctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgcagtcctctac 8121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7739 CTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGTCCTCTAC 7798

Qy   8122 ccctgcgtgg 8131
          ||||||||||
Db   7799 CCCTGCGTGG 7808
```

```
RESULT   8
AF038601
LOCUS       AF038601     7333 bp    mRNA            MAM       21-MAY-1998
DEFINITION  Sus scrofa porcine endogenous retrovirus ERV-PK15 mRNA, complete
            sequence.
ACCESSION   AF038601
VERSION     AF038601.1  GI:3133304
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 7333)
  AUTHORS   Akiyoshi,D.E., Denaro,M., Zhu,H., Greenstein,J.L., Banerjee,P. and
            Fishman,J.A.
  TITLE     Identification of a full-length cDNA for an endogenous retrovirus
            of miniature swine
  JOURNAL   J. Virol. 72 (5), 4503-4507 (1998)
  MEDLINE   98216827
REFERENCE   2  (bases 1 to 7333)
  AUTHORS   Fishman,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1997) Infectious Disease Unit, Massachusetts
            General Hospital, 55 Fruit Street, Boston, MA 02114, USA
REFERENCE   3  (bases 1 to 7333)
  AUTHORS   Fishman,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-1998) Infectious Disease Unit, Massachusetts
            General Hospital, 55 Fruit Street, Boston, MA 02114, USA
  REMARK    Sequence update by submitter
```

```
COMMENT     On May 16, 1998 this sequence version replaced gi:3116445.
FEATURES             Location/Qualifiers
     source          1. .7333
                     /organism="Sus scrofa"
                     /strain="miniature swine"
                     /db_xref="taxon:9823"
                     /cell_line="PK-15"
                     /note="porcine endogenous retrovirus ERV-PK15; infective
                     virus but represents a naturally occurring transcript"
     LTR             1. .578
     CDS             598. .2172
                     /codon_start=1
                     /product="gag protein"
                     /protein_id="AAC16766.1"
                     /db_xref="GI:3116446"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPGSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPRIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     KKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREEREERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKD"
     CDS             2320. .3552
                     /note="truncated protein"
                     /codon_start=1
                     /product="pol protein"
                     /protein_id="AAC16767.1"
                     /db_xref="GI:3133305"
                     /translation="MGATGQRQYPWTTRRTVDLGVGRVTHSFLVIPECPVPLLGRDLL
                     TKMGAQISFEQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAW
                     AETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQ
                     SPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVL
                     DLKDAFFCLRLHPTSQPLFTFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDL
                     ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRRE
                     VTYLGYSLRGGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAP
                     LYPLTKEKGGCLPQQGGK"
     CDS             4803. .6725
                     /note="env protein; N terminus deleted"
                     /codon_start=1
                     /pseudo
     LTR             6726. .7333
BASE COUNT     1984 a   1788 c   1885 g   1676 t
ORIGIN
```

Query Match 36.1%; Score 2967; DB 4; Length 7333;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3507; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

```
Qy       1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
Db      11 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 70

Qy      61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgagggggat 120
Db      71 AAGACCGCTTCTTGTGAGTGATTTGGGGTGTCGCCTCTTCCGAGCCCGGACGAGGGGGAT 130

Qy     121 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctta 180
Db     131 TGTTCTTTTACTGGCCTTTCATTTGGTGCGTTGGCCGGGAAATCCTGCGACCACCCCTTA 190

Qy     181 cacccgagaaccgacttggaggtaaagggatccccttggaacgtgtgtgtgtgtcggcc 240
Db     191 CACCCGAGAACCGACTTGGAGGTAAAGGGATCCCCTTTGGAACATATGTGTGTGTCGGCC 250

Qy     241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
Db     251 GGCGTCTCTGTTCTGAGTGTCTGTTTTCGGTGATGCGCGCTTTCGGTTTGCAGCTGTCCT 310

Qy     301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
Db     311 CTCAGACCGTAAGGACTGGAGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAGGCTG 370

Qy     361 ccaccctgggggacgccccgggaggtgggggagagccagggacgcctggtggtctcctact 420
Db     371 CCACCCTGGGGGACGCCCCGGGAGGTGGGGAGAGCCAGGGACGCCTGGTGGTCTCCTACT 430

Qy     421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 480
Db     431 GTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCG 490

Qy     481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
Db     491 ACTCTTTTGCCTGCTTGTGGAAGACGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGTTT 550

Qy     541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
Db     551 CTGTTTCGTGTGTCTTTGTCTTGTGCGTCCTTGTCTACAGTTTTAATATGGGACAGACAG 610

Qy     601 tgactacccccctttagtttgactctcgaccattggactgaagttagatccagggctcata 660
Db     611 TGACTACCCCCCTTAGTTTGACTCTCGACCATTGGACTGAAGTTAGATCCAGGGCTCATA 670

Qy     661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
Db     671 ATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGCCTCTGAATGGCCAA 730

Qy     721 cattcgatgttggatggccatcagagggaccttttaattctgaaattatcctggctgtta 780
Db     731 CATTCGATGTTGGATGGCCATCAGAGGGACCTTTAATTCTGAAATTATCCTGGCTGTTA 790

Qy     781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 840
Db     791 AGGCAATCATTTTTCAGACTGGACCCGGCTCTCATCCTGATCAGGAGCCCTATATCCTTA 850

Qy     841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
Db     851 CGTGGCAAGATTTGGCAGAAGATCCTCCGCCATGGGTTAAACCATGGCTAAATAAACCAA 910

Qy     901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
Db     911 GAAAGCCAGGTCCCCGAATCCTGGCTCTTGGAGAGAAAAACAAACACTCGGCCGAAAAAG 970

Qy     961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
Db     971 TCGAGCCCTC-TCCTCGTATCTACCCCGAGATCGAGGAGCCGCCGACTTGGCCGGAACCC 1029

Qy    1021 caacctgttccccccaccccctttatccagcacagggtgctgtgagggga-cctctgccccct 1079
Db    1030 CAACCTGTTCCCCCACCCCCTTATCCAGCACAGGGTGCTGTGAGGGGACCCTCTGCCCCT 1089

Qy    1080 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1139
Db    1090 CCTGGAGCTCCGGTGGTGGAGGGACCTGCTGCCGGGACTCGGAGCCGGAGAGGCGCCACC 1149

Qy    1140 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1199
Db    1150 CCGGAGCGGACAGACGAGATCGCGATATTACCGCTGCGCACCTATGGCCCTCCCATGCCA 1209

Qy    1200 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1259
Db    1210 GGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGG 1269

Qy    1260 aaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtcc 1319
Db    1270 AAAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGAGTCC 1329

Qy    1320 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1379
Db    1330 CTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTC 1389

Qy    1380 acaaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgac 1439
```

```
Db   1390 ACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAAAAAAAATGTTCCTGGGGCCGAC 1449

Qy   1440 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1450 GGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGATTTCCCTTGACTCGCCCCGGT 1509

Qy   1500 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1510 TGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCTGGTG 1569

Qy   1560 gcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatg 1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1570 GCGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAGGTGATG 1629

Qy   1620 cagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcgg 1679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1630 CAGGGACCGAACGAACCTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGG 1689

Qy   1680 ttcacccctttgatcctacctcagaggcccagaaagcctcagtggccctggccttcatt 1739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1690 TTCACCCCTTTGATCCTACCTCAGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTCATT 1749

Qy   1740 gggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggct 1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1750 GGGCAGTCGGCTCTGGATATCAGGAAGAAACTTCAGAGACTGGAAGGGTTACAGGAGGCT 1809

Qy   1800 gagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagag 1859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1810 GAGTTACGTGATCTAGTGAGAGAGGCAGAGAAGGTGTATTACAGAAGGGAGACAGAAGAG 1869

Qy   1860 gagaaggaacagagaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacgg 1919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1870 GAGAAGGAACAGAGAAAAGAAAAGGAGAGAGAAGAAAGGGAGGAAAGACGTGATAGACGG 1929

Qy   1920 caagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggag 1979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1930 CAAGAGAAGAATTTGACTAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAG 1989

Qy   1980 agagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaat 2039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1990 AGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAAT 2049

Qy   2040 aggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagg 2099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2050 AGGACCCCACTCGACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGGACACTGGGCAAGG 2109

Qy   2100 aactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagat 2159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2110 AACTGCCCCAAGAAGGGAAACAAAGGACCGAAGGTCCTAGCTCTAGAAGAAGATAAAGAT 2169

Qy   2160 taggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtggagggg 2219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2170 TAGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGAGGGG 2229

Qy   2220 caaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccatta 2279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2230 CAACCAGTTGAGTTCCTGGTTGATACCGGAGCGGAGCATTCAGTGCTGCTACAACCATTA 2289

Qy   2280 ggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatcca 2339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2290 GGAAAACTAAAAGAAAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACGGCAGTATCCA 2349

Qy   2340 tggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtc 2399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2350 TGGACTACCCGAAGAACCGTTGACTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTC 2409

Qy   2400 atccctgagtgcccagtacccctctaggtagagacttactgaccaagatgggagctcaa 2459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2410 ATCCCTGAGTGCCCAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAA 2469

Qy   2460 atttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttg 2519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2470 ATTTCTTTTGAACAAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCACTGTGTTG 2529

Qy   2520 accctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagat 2579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2530 ACCCTCCAATTAGATGATGAATATCGACTATATTCTCCCCAAGTAAAGCCTGATCAAGAT 2589

Qy   2580 atacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttg 2639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2590 ATACAGTCCTGGTTGGAGCAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGGTTTG 2649

Qy   2640 gcaaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtc 2699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2650 GCAAAGCAAGTTCCCCCACAGGTTATTCAACTGAAGGCCAGTGCTACACCAGTATCAGTC 2709

Qy   2700 agacagtacccctgagtagagaggctcgagaaggaatttggccgcatgttcaaagatta 2759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2710 AGACAGTACCCCTTGAGTAGAGAGGCTCGAGAAGGAATTTGGCCGCATGTTCAAAGATTA 2769

Qy   2760 atccaacagggcatcctagttcctgtccaatccccttggaatactccctgctaccggtt 2819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2770 ATCCAACAGGGCATCCTAGTTCCTGTCCAATCCCCTTGGAATACTCCCCTGCTACCGGTT 2829

Qy   2820 aggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaagg 2879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2830 AGGAAGCCTGGGACCAATGATTATCGACCAGTACAGGACTTGAGAGAGGTCAATAAAAGG 2889

Qy   2880 gtgcaggacatacacccaacggtcccgaacccttataacctcttgagcgccctcccgcct 2939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2890 GTGCAGGACATACACCCAACGGTCCCGAACCCTTATAACCTCTTGAGCGCCCTCCCGCCT 2949

Qy   2940 gaacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacac 2999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2950 GAACGGAACTGGTACACAGTATTGGACTTAAAAGATGCCTTCTTCTGCCTGAGATTACAC 3009

Qy   3000 cccactagccaaccacttttttgccttcgaatggagagatccaggtacgggaagaaccggg 3059
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3010 CCCACTAGCCAACCACTTTTTACCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGG 3069

Qy   3060 cagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaa 3119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3070 CAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGACGAA 3129

Qy   3120 gccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccag 3179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3130 GCCCTACACAGGGACCTGGCCAACTTCAGGATCCAACACCCTCAGGTGACCCTCCTCCAG 3189

Qy   3180 tacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaag 3239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3190 TACGTGGATGACCTGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGTACGAAG 3249

Qy   3240 gcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagatt 3299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3250 GCACTACTGCTGGAATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATT 3309

Qy   3300 tgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacg 3359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3310 TGCAGGAGAGAGGTAACATACTTGGGGTACAGTTTGCGGGGCGGGCAGCGATGGCTGACG 3369

Qy   3360 gaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgaga 3419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3370 GAGGCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAGTGAGA 3429

Qy   3420 gagtttttggggacagctggatttttgcagactgtggatcccggggtttgcgaccttagca 3479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3430 GAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTTAGCA 3489

Qy   3480 gccccactctacccgctaaccaaagaaaaaggggg 3514
          ||||||||||||||||||||||||||||||||||
Db   3490 GCCCCACTCTACCCGCTAACCAAAGAAAAAGGGGG 3524


RESULT   9
AR130474
LOCUS       AR130474      7333 bp     DNA             PAT       16-MAY-2001
```

```
DEFINITION  Sequence 2 from patent US 6190861.
ACCESSION   AR130474
VERSION     AR130474.1  GI:14118799
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7333)
  AUTHORS   Fishman,J.A.
  TITLE     Molecular sequences of swine retroviruses method of using
  JOURNAL   Patent: US 6190861-A 2 20-FEB-2001;
FEATURES             Location/Qualifiers
     source          1. .7333
                     /organism="unknown"
BASE COUNT     1984 a   1788 c   1885 g   1676 t
ORIGIN
```

Query Match 36.1%; Score 2967; DB 6; Length 7333;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3507; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

```
Qy       1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 70

Qy      61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggggat 120
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 AAGACCGCTTCTTGTGAGTGATTTGGGGTGTCGCCTCTTCCGAGCCCGGACGAGGGGGAT 130

Qy     121 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctta 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     131 TGTTCTTTTACTGGCCTTTCATTTGGTGCGTTGGCCGGGAAATCCTGCGACCACCCCTTA 190

Qy     181 cacccgagaaccgacttggaggtaaagggatcccctttggaacgtgtgtgtgtgtcggcc 240
           |||||||||||||||||||||||||||||||||||| | ||||||||||||||||||||
Db     191 CACCCGAGAACCGACTTGGAGGTAAAGGGATCCCCTTTGGAACATATGTGTGTGTCGGCC 250

Qy     241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 GGCGTCTCTGTTCTGAGTGTCTGTTTTCGGTGATGCGCGCTTTCGGTTTGCAGCTGTCCT 310

Qy     301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 CTCAGACCGTAAGGACTGGAGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAGGCTG 370

Qy     361 ccacccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     371 CCACCCTGGGGGACGCCCCGGGAGGTGGGGAGAGCCAGGGACGCCTGGTGGTCTCCTACT 430

Qy     421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     431 GTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCG 490

Qy     481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     491 ACTCTTTTGCCTGCTTGTGGAAGACGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGTTT 550

Qy     541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     551 CTGTTTCGTGTGTCTTTGTCTTGTGCGTCCTTGTCTACAGTTTTAATATGGGACAGACAG 610

Qy     601 tgactacccccctttagtttgactctcgaccattggactgaagttagatccagggctcata 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     611 TGACTACCCCCCTTAGTTTGACTCTCGACCATTGGACTGAAGTTAGATCCAGGGCTCATA 670

Qy     661 atttgtcagttcaggttaagaagggaccttggcagacttttctgtgcctctgaatggccaa 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     671 ATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGCCTCTGAATGGCCAA 730

Qy     721 cattcgatgttggatggccatcagaggggaccttttaattctgaaattatcctggctgtta 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     731 CATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCTGAAATTATCCTGGCTGTTA 790

Qy     781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     791 AGGCAATCATTTTTCAGACTGGACCCGGCTCTCATCCTGATCAGGAGCCCTATATCCTTA 850

Qy     841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     851 CGTGGCAAGATTTGGCAGAAGATCCTCCGCCATGGGTTAAACCATGGCTAAATAAACCAA 910

Qy     901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     911 GAAAGCCAGGTCCCCGAATCCTGGCTCTTGGAGAGAAAAACAAACACTCGGCCGAAAAAG 970

Qy     961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     971 TCGAGCCCTC-TCCTCGTATCTACCCCGAGATCGAGGAGCCGCCGACTTGGCCGGAACCC 1029

Qy    1021 caacctgttcccccacccccttatccagcacagggtgctgtgagggga-cctctgcccct 1079
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    1030 CAACCTGTTCCCCCACCCCCTTATCCAGCACAGGGTGCTGTGAGGGGACCCTCTGCCCCT 1089

Qy    1080 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1139
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1090 CCTGGAGCTCCGGTGGTGGAGGGACCTGCTGCCGGGACTCGGAGCCGGAGAGGCGCCACC 1149

Qy    1140 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1199
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1150 CCGGAGCGGACAGACGAGATCGCGATATTACCGCTGCGCACCTATGGCCCTCCCATGCCA 1209

Qy    1200 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1210 GGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGG 1269

Qy    1260 aaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtcc 1319
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1270 AAAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGAGTCC 1329

Qy    1320 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1379
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1330 CTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTC 1389

Qy    1380 acaaccgaggagcgagagagaagaattctgttagaggctagaaaaaaatgttcctggggccgac 1439
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    1390 ACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAAAAAAAATGTTCCTGGGGCCGAC 1449

Qy    1440 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1499
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1450 GGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGATTTCCCTTGACTCGCCCCGGT 1509

Qy    1500 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1559
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1510 TGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCTGGTG 1569

Qy    1560 gcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatg 1619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1570 GCGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAGGTGATG 1629

Qy    1620 cagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcgg 1679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1630 CAGGGACCGAACGAACCTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGG 1689

Qy    1680 ttcaccccttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcatt 1739
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1690 TTCACCCCTTTTGATCCTACCTCAGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTCATT 1749

Qy    1740 gggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggct 1799
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1750 GGGCAGTCGGCTCTGGATATCAGGAAGAAACTTCAGAGACTGGAAGGGTTACAGGAGGCT 1809

Qy    1800 gagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagag 1859
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1810 GAGTTACGTGATCTAGTGAGAGAGGCAGAGAAGGTGTATTACAGAAGGGAGACAGAAGAG 1869
```

```
Qy   1860 gagaaggaacagagaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacgg 1919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1870 GAGAAGGAACAGAGAAAAGAAAAGGAGAGAGAAGAAAGGGAGGAAAGACGTGATAGACGG 1929

Qy   1920 caagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggag 1979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1930 CAAGAGAAGAATTTGACTAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAG 1989

Qy   1980 agagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaat 2039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1990 AGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAAT 2049

Qy   2040 aggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagg 2099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2050 AGGACCCCACTCGACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGGACACTGGGCAAGG 2109

Qy   2100 aactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagat 2159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2110 AACTGCCCCAAGAAGGGAAACAAAGGACCGAAGGTCCTAGCTCTAGAAGAAGATAAAGAT 2169

Qy   2160 taggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtggagggg 2219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2170 TAGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGAGGGG 2229

Qy   2220 caaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccatta 2279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2230 CAACCAGTTGAGTTCCTGGTTGATACCGGAGCGGAGCATTCAGTGCTGCTACAACCATTA 2289

Qy   2280 ggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatcca 2339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2290 GGAAAACTAAAAGAAAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACGGCAGTATCCA 2349

Qy   2340 tggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtc 2399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2350 TGGACTACCCGAAGAACCGTTGACTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTC 2409

Qy   2400 atccctgagtgcccagtacccctttctaggtagagacttactgaccaagatgggagctcaa 2459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2410 ATCCCTGAGTGCCCAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAA 2469

Qy   2460 atttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttg 2519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2470 ATTTCTTTTGAACAAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCACTGTGTTG 2529

Qy   2520 accctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagat 2579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2530 ACCCTCCAATTAGATGATGAATATCGACTATATTCTCCCCAAGTAAAGCCTGATCAAGAT 2589

Qy   2580 atacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttg 2639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2590 ATACAGTCCTGGTTGGAGCAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGGTTTG 2649

Qy   2640 gcaaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtc 2699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2650 GCAAAGCAAGTTCCCCCACAGGTTATTCAACTGAAGGCCAGTGCTACACCAGTATCAGTC 2709

Qy   2700 agacagtacccccttgagtagagaggctcgagaaggaatttggccgcatgttcaaagatta 2759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2710 AGACAGTACCCCTTGAGTAGAGAGGCTCGAGAAGGAATTTGGCCGCATGTTCAAAGATTA 2769

Qy   2760 atccaacagggcatcctagttcctgtccaatccccttggaatactccccctgctaccggtt 2819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2770 ATCCAACAGGGCATCCTAGTTCCTGTCCAATCCCCTTGGAATACTCCCCTGCTACCGGTT 2829

Qy   2820 aggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaagg 2879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2830 AGGAAGCCTGGGACCAATGATTATCGACCAGTACAGGACTTGAGAGAGGTCAATAAAAGG 2889

Qy   2880 gtgcaggacatacacccaacggtcccgaacccttataacctcttgagcgccctcccgcct 2939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2890 GTGCAGGACATACACCCAACGGTCCCGAACCCTTATAACCTCTTGAGCGCCCTCCCGCCT 2949

Qy   2940 gaacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacac 2999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2950 GAACGGAACTGGTACACAGTATTGGACTTAAAAGATGCCTTCTTCTGCCTGAGATTACAC 3009

Qy   3000 cccactagccaaccacttttttgccttcgaatggagagatccaggtacgggaagaaccggg 3059
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3010 CCCACTAGCCAACCACTTTTTACCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGG 3069

Qy   3060 cagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaa 3119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3070 CAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGACGAA 3129

Qy   3120 gccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccag 3179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3130 GCCCTACACAGGGACCTGGCCAACTTCAGGATCCAACACCCTCAGGTGACCCTCCTCCAG 3189

Qy   3180 tacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaag 3239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3190 TACGTGGATGACCTGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGTACGAAG 3249

Qy   3240 gcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagatt 3299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3250 GCACTACTGCTGGAATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATT 3309

Qy   3300 tgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacg 3359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3310 TGCAGGAGAGAGGTAACATACTTGGGGTACAGTTTGCGGGGCGGGCAGCGATGGCTGACG 3369

Qy   3360 gaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgaga 3419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3370 GAGGCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAGTGAGA 3429

Qy   3420 gagtttttggggacagctggattttgcagactgtggatcccggggtttgcgaccttagca 3479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3430 GAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTTAGCA 3489

Qy   3480 gccccactctacccgctaaccaaagaaaaaggggg 3514
          ||||||||||||||||||||||||||||||||||
Db   3490 GCCCCACTCTACCCGCTAACCAAAGAAAAAGGGGG 3524
```

RESULT 10
A66551
LOCUS A66551 3320 bp DNA PAT 29-MAR-1999
DEFINITION Sequence 1 from Patent WO9740167.
ACCESSION A66551
VERSION A66551.1 GI:4538104
KEYWORDS .
SOURCE porcine endogenous retrovirus.
ORGANISM porcine endogenous retrovirus
Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses; 1-Mammalian type C virus group.
REFERENCE 1 (bases 1 to 3320)
AUTHORS Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
TITLE PORCINE RETROVIRUS
JOURNAL Patent: WO 9740167-A 1 30-OCT-1997;
Q ONE BIOTECH LTD (GB)
FEATURES Location/Qualifiers
source 1. .3320
/organism="porcine endogenous retrovirus"
/db_xref="taxon:61673"
BASE COUNT 922 a 844 c 832 g 722 t
ORIGIN

Query Match 32.7%; Score 2682; DB 6; Length 3320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   2972 agatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcgaatg 3031
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 AGATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAACCACTTTTTGCCTTCGAATG 80

Qy   3032 gagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggtt 3091
```

```
Db      81 GAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTT 140

Qy    3092 caagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggat 3151
Db     141 CAAGAACTCCCCGACCATCTTTGACGAAGCCCTACACAGGGACCTGGCCAACTTCAGGAT 200

Qy    3152 ccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccac 3211
Db     201 CCAACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCAC 260

Qy    3212 caaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggcta 3271
Db     261 CAAACAGGACTGCTTAGAAGGTACGAAGGCACTACTGCTGGAATTGTCTGACCTAGGCTA 320

Qy    3272 cagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacag 3331
Db     321 CAGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGGTAACATACTTGGGGTACAG 380

Qy    3332 tttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagatacc 3391
Db     381 TTTGCGGGGCGGGCAGCGATGGCTGACGGAGGCACGGAAGAAAACTGTAGTCCAGATACC 440

Qy    3392 ggccccaaccacagccaaacaagtgagagagtttttggggacagctggatttttgcagact 3451
Db     441 GGCCCCAACCACAGCCAAACAAGTGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACT 500

Qy    3452 gtggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaagg 3511
Db     501 GTGGATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGG 560

Qy    3512 gggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgct 3571
Db     561 GGGATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATGCTATCAAAAAGGCCCTGCT 620

Qy    3572 gagcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatga 3631
Db     621 GAGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCTTTACCCTTTATGTGGATGA 680

Qy    3632 gcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgt 3691
Db     681 GCGTAAGGGAGTAGCCCGAGGAGTTTTAACCCAAACCCTAGGACCATGGAGGAGACCTGT 740

Qy    3692 tgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggc 3751
Db     741 TGCCTACCTGTCAAAGAAGCTTGATCCTGTAGCCAGTGGTTGGCCCGTATGTCTGAAGGC 800

Qy    3752 tatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatat 3811
Db     801 TATCGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAGAATAT 860

Qy    3812 aactgtaatagcccccatgcattggagaacatcgttcggcagccccagaccgatggat 3871
Db     861 AACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAGACCGATGGAT 920

Qy    3872 gaccaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgc 3931
Db     921 GACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCACAGAGAGGGTCACTTTCGC 980

Qy    3932 tccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgac 3991
Db     981 TCCACCAGCCGCTCTCAACCCTGCCACTCTTCTGCCTGAAGAGACTGATGAACCAGTGAC 1040

Qy    3992 tcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacat 4051
Db    1041 TCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACAT 1100

Qy    4052 accgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaagg 4111
Db    1101 ACCGCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGG 1160

Qy    4112 taagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcct 4171

Db    1161 TAAGAGGATGGCTGGGGCGGCAGTGGTGGACGGGACCCGCACGATCTGGGCCAGCAGCCT 1220

Qy    4172 gccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggct 4231
Db    1221 GCCGGAAGGAACTTCAGCGCAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCT 1280

Qy    4232 ggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcaca 4291
Db    1281 GGCCGAAGGGAAATCCATAAACATTTATACGGACAGCAGGTATGCCTTTGCGACTGCACA 1340

Qy    4292 cgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaa 4351
Db    1341 CGTACACGGGGCCATCTATAAACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAA 1400

Qy    4352 gaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctat 4411
Db    1401 GAACAAAGAGGAAATTCTAAGCCTATTAGAAGCCTTACATTTGCCAAAAAGGCTAGCTAT 1460

Qy    4412 tatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggc 4471
Db    1461 TATACACTGTCCTGGACATCAGAAAGCCAAAGATCTCATATCTAGAGGGAACCAGATGGC 1520

Qy    4472 tgaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcc 4531
Db    1521 TGACCGGGTTGCCAAGCAGGCAGCCCAGGCTGTTAACCTTCTGCCTATAATAGAAACGCC 1580

Qy    4532 caaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagat 4591
Db    1581 CAAAGCCCCAGAACCCAGACGACAGTACACCCTAGAAGACTGGCAAGAGATAAAAAAGAT 1640

Qy    4592 agaccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaatcct 4651
Db    1641 AGACCAGTTCTCTGAGACTCCGGAGGGGACCTGCTATACCTCATATGGGAAGGAAATCCT 1700

Qy    4652 gccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaac 4711
Db    1701 GCCCCACAAAGAAGGGTTAGAATATGTCCAACAGATACATCGTCTAACCCACCTAGGAAC 1760

Qy    4712 taaacacctgcagcagttggtcagaacatcccccttatcatgttctgaggctaccaggagt 4771
Db    1761 TAAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGGCTACCAGGAGT 1820

Qy    4772 ggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccag 4831
Db    1821 GGCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGGTTAATGCTAATCCTTCCAG 1880

Qy    4832 aatacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggactt 4891
Db    1881 AATACCTCCAGGAAAGAGACTAAGGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTT 1940

Qy    4892 cactgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctt 4951
Db    1941 CACTGAGGTAAAGCCGGCTAAATACGGAAACAAATATCTATTGGTTTTTGTAGACACCTT 2000

Qy    4952 ttcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaa 5011
Db    2001 TTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAGAA 2060

Qy    5012 aatactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatgg 5071
Db    2061 AATACTGGAGGAAATTTTTCCAAGATTTGGAATACCTAAGGTAATAGGGTCAGACAATGG 2120

Qy    5072 tccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaa 5131
Db    2121 TCCAGCTTTCGTTGCCCAGGTAAGTCAGGGACTGGCCAAGATATTGGGGATTGATTGGAA 2180

Qy    5132 actgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccat 5191
Db    2181 ACTGCATTGTGCATACAGACCCCAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCAT 2240

Qy    5192 taaagagacccttaccaaattgaccacagagactggcattaatgattggatggctctcct 5251
Db    2241 TAAAGAGACCCTTACCAAATTGACCACAGAGACTGGCATTAATGATTGGATGGCTCTCCT 2300
```

```
Qy   5252 gccctttgtgctttttagggtgaggaacacccctggacagtttgggctgacccccatga 5311
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2301 GCCCTTTGTGCTTTTTAGGGTGAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGA 2360

Qy   5312 attgctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctgatgt 5371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2361 ATTGCTCTACGGGGGACCCCCCCCCGTTGGCAGAAATTGCCTTTGCACATAGTGCTGATGT 2420

Qy   5372 gctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcgagc 5431
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2421 GCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAGCGAGC 2480

Qy   5432 gtggaagcagctccgggaggcctactcaggaggagacttgcaagttccacatcgcttcca 5491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2481 GTGGAAGCAGCTCCGGGAGGCCTACTCAGGAGGAGACTTGCAAGTTCCACATCGCTTCCA 2540

Qy   5492 agttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaa 5551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2541 AGTTGGAGATTCAGTCTATGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAA 2600

Qy   5552 gggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatccccac 5611
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2601 GGGACCTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCCCCAC 2660

Qy   5612 ctggatccatgcatcccacgttaagccggcgccacctcccga 5653
          |||||||||||||||||||||||||||||||||||||||||
Db   2661 CTGGATCCATGCATCCCACGTTAAGCCGGCGCCACCTCCCGA 2702


RESULT  11
AF147808
LOCUS       AF147808     3025 bp    DNA             MAM       27-APR-2000
DEFINITION  Sus scrofa domestica porcine endogenous retrovirus type C gag-pol
            precursor, gene, partial cds.
ACCESSION   AF147808
VERSION     AF147808.1  GI:7650139
KEYWORDS    .
SOURCE      domestic pig.
  ORGANISM  Sus scrofa domestica
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 3025)
  AUTHORS   Blusch,J.H., Seelmeir,S. and von der Helm,K.
  TITLE     Evidence for extreme sequence homogeneity of the PERV A/B 5'
            LTR/leader/gag/prot region by predicted targeted amplification from
            a Du X DL pig
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3025)
  AUTHORS   Blusch,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-1999) Virology, Max-von-Pettenkofer-Institute,
            Pettenkofer-Str. 9a, Munich 80336, Germany
FEATURES             Location/Qualifiers
     source          1. .3025
                     /organism="Sus scrofa domestica"
                     /strain="breed Duroc x German Landbreed, Du x DL3791"
                     /sub_species="domestica"
                     /db_xref="taxon:9825"
     repeat_region   1. .3025
                     /note="PERV A/B; porcine endogenous retrovirus type C"
                     /rpt_family="perv"
                     /rpt_type=dispersed
     primer_bind     587. .605
                     /note="primer binding site for tRNA-Gly"
     CDS             1032. .>3025
                     /note="gag protein produced by proteolytic cleavage of the
                     mature protease out of the precursor"
                     /codon_start=1
                     /transl_except=(pos:2604. .2606,aa:Gln)
                     /product="gag-pol precursor"
                     /protein_id="AAF65926.1"
                     /db_xref="GI:7650140"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPGSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPRIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREREEERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKDQGRRGSDPLPEPRVTLKVEGQPVEFLVDTGAEHSVLLQPLGK
                     LKEKKSWVMGATGQRQYPWTTRRTVDLGVGRVTHSFLVIPECPVPLLGRDLLTKMGAQ
                     ISFEQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQD"
     mat_peptide     1032. .2603
                     /product="gag protein"
     mat_peptide     2604. .>3025
                     /product="aspartic protease"
BASE COUNT      743 a    749 c    849 g    684 t
ORIGIN


  Query Match             24.8%;  Score 2034;  DB 4;  Length 3025;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 2574;  Conservative    0;  Mismatches    6;  Indels    2;  Gaps    2;

Qy      1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    445 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 504

Qy     61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505 AAGACCGCTTCTCGTGAGTGATTTGGGGTGTCGCCTCTTCCGAGCCCGGACGAGGGGGAT 564

Qy    121 tgttcttttactggccttttcatttggtgcgttggccgggaaatcctgcgaccaccccctta 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    565 TGTTCTTTTACTGGCCTTTCATTTGGTGCGTTGGCCGGGAAATCCTGCGACCACCCCTTA 624

Qy    181 cacccgagaaccgacttggaggtaaagggatccccttggaacgtgtgtgtgtgtcggcc 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    625 CACCCGAGAACCGACTTGGAGGTAAAGGGATCCCCTTTGGAAAGTGTGTGTGTCGGCC 684

Qy    241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    685 GGCGTCTCTGTTCTGAGTGTCTGTTTTCGGTGATGCGCGCTTTCGGTTTGCAGCTGTCCT 744

Qy    301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    745 CTCAGACCGTAAGGACTGGAGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAGGCTG 804

Qy    361 ccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    805 CCACCCTGGGGGACGCCCCGGGAGGTGGGGAGAGCCAGGGACGCCTGGTGGTCTCCTACT 864

Qy    421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    865 GTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCG 924

Qy    481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    925 ACTCTTTTGCCTGCTTGTGGAAGACGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGTTT 984

Qy    541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    985 CTGTTTCGTGTGTCTTTGTCTTGTGCGTCCTTGTCTACAGTTTTAATATGGGACAGACAG 1044

Qy    601 tgactacccccctagtttgactctcgaccattggactgaagttagatccagggctcata 660
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1045 TGACGACCCCCCTTAGTTTGACTCTCGACCATTGGACTGAAGTTAGATCCAGGGCTCATA 1104

Qy    661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1105 ATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGCCTCTGAATGGCCAA 1164

Qy    721 cattcgatgttggatggccatcagaggggacctttaattctgaaattatcctggctgtta 780
```

```
Db  1165 CATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCTGAAATTATCCTGGCTGTTA 1224

Qy   781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatccttta 840

Db  1225 AGGCAATCATTTTTCAGACTGGACCCGGCTCTCATCCTGATCAGGAGCCCTATATCCTTA 1284

Qy   841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900

Db  1285 CGTGGCAAGATTTGGCAGAAGATCCTCCGCCATGGGTTAAACCATGGCTAAATAAACCAA 1344

Qy   901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960

Db  1345 GAAAGCCAGGTCCCCGAATCCTGGCTCTTGGAGAGAAAAACAAACACTCGGCCGAAAAAG 1404

Qy   961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020

Db  1405 TCGAGCCCTC-TCCTCGTATCTACCCCGAGATCGAGGAGCCGCCGACTTGGCCGGAACCC 1463

Qy  1021 caacctgttcccccacccccttatccagcacagggtgctgtgaggga-cctctgcccct 1079

Db  1464 CAACCTGTTCCCCCACCCCCTTATCCAGCACAGGGTGCTGTGAGGGGACCCTCTGCCCCT 1523

Qy  1080 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1139

Db  1524 CCTGGAGCTCCGGTGGTGGAGGGACCTGCTGCCGGGACTCGGAGCCGGAGAGGCGCCACC 1583

Qy  1140 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1199

Db  1584 CCGGAGCGGACAGACGAGATCGCGATATTACCGCTGCGCACCTATGGCCCTCCCATGCCG 1643

Qy  1200 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1259

Db  1644 GGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGG 1703

Qy  1260 aaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtcc 1319

Db  1704 AAAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGAGTCC 1763

Qy  1320 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1379

Db  1764 CTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTC 1823

Qy  1380 acaaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgac 1439

Db  1824 ACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAGAAAAAATGTTCCTGGGGCCGAC 1883

Qy  1440 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1499

Db  1884 GGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGATTTCCCTTGACTCGCCCCGGT 1943

Qy  1500 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1559

Db  1944 TGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCTGGTG 2003

Qy  1560 gcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatg 1619

Db  2004 GCGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAGGTGATG 2063

Qy  1620 cagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcgg 1679

Db  2064 CAGGGACCGAACGAACCTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGG 2123

Qy  1680 ttcaccccttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcatt 1739

Db  2124 TTCACCCCTTTTGATCCTACCTCGGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTCATT 2183

Qy  1740 gggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggct 1799

Db  2184 GGGCAGTCGGCTCTGGATATCAGAAAGAAACTTCAGAGACTGGAAGGGTTACAGGAGGCT 2243

Qy  1800 gagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagag 1859

Db  2244 GAGTTACGTGATCTAGTGAGAGAGGCAGAGAAGGTGTATTACAGAAGGGAGACAGAAGAG 2303

Qy  1860 gagaaggaacagagaaaagaaaaggagagagaagaaaggggaggaaagacgtgatagacgg 1919

Db  2304 GAGAAGGAACAGAGAAAAGAAAAGGAGAGAGAAGAAAGGGAGGAAAGACGTGATAGACGG 2363

Qy  1920 caagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggag 1979

Db  2364 CAAGAGAAGAATTTGACTAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAG 2423

Qy  1980 agagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaat 2039

Db  2424 AGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAAT 2483

Qy  2040 aggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagg 2099

Db  2484 AGGACCCCACTCGACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGGACACTGGGCAAGG 2543

Qy  2100 aactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagat 2159

Db  2544 AACTGCCCCAAGAAGGGAAACAAAGGACCGAAGGTCCTAGCTCTAGAAGAAGATAAAGAT 2603

Qy  2160 taggggagacggggttcggaccccctccccgagcccagggtaactttgaaggtggagggg 2219

Db  2604 TAGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGAGGGG 2663

Qy  2220 caaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccatta 2279

Db  2664 CAACCAGTTGAGTTCCTGGTTGATACCGGAGCGGAGCATTCAGTGCTGCTACAACCATTA 2723

Qy  2280 ggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatcca 2339

Db  2724 GGAAAACTAAAAGAAAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACGGCAGTATCCA 2783

Qy  2340 tggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtc 2399

Db  2784 TGGACTACCCGAAGAACCGTTGACTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTC 2843

Qy  2400 atccctgagtgcccagtacccctctaggtagagacttactgaccaagatgggagctcaa 2459

Db  2844 ATCCCTGAGTGCCCAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAA 2903

Qy  2460 atttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttg 2519

Db  2904 ATTTCTTTTGAACAAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCACTGTGTTG 2963

Qy  2520 accctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagat 2579

Db  2964 ACCCTCCAATTAGATGATGAATATCGACTATATTCTCCCCAAGTAAAGCCTGATCAAGAT 3023

Qy  2580 at 2581
         ||
Db  3024 AT 3025
```

```
RESULT  12
AF014162
LOCUS       AF014162     1971 bp    mRNA            VRL       01-JUN-1998
DEFINITION  Porcine endogenous retrovirus envelope protein (env) mRNA, complete
            cds.
ACCESSION   AF014162
VERSION     AF014162.1  GI:3169735
KEYWORDS    .
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 1971)
  AUTHORS   Haworth,C., Galbraith,D.N., Lees,G.M. and Smith,K.T.
  TITLE     Porcine endogenous retrovirus (PoEV) Env sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1971)
  AUTHORS   Haworth,C., Galbraith,D.N., Lees,G.M. and Smith,K.T.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1997) Q-One Biotech Ltd., Todd Campus, Acre Rd.,
            Glasgow G20 0XA, UK
FEATURES             Location/Qualifiers
     source          1. .1971
                     /organism="porcine endogenous retrovirus"
                     /db_xref="taxon:61673"
                     /cell_line="derived from cell line PK15"
                     /note="PoEV"
     gene            1. .1971
                     /gene="env"
     CDS             1. .1971
                     /gene="env"
                     /codon_start=1
                     /product="envelope protein"
                     /protein_id="AAC17948.1"
                     /db_xref="GI:3169736"
                     /translation="MHPTLSRRHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSK
                     RLIDSSNPHRPLSLTWLIIDPDTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTP
                     PNLVRSYGFYCCPGTEKEKYCGGSGESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVN
                     SGPGKYKMMKLYKDKSCSPSDLDYLKISFTERKTGKYSKVDKWYELGNSFLLYGGGAG
                     STLTIRLRIETGTEPPVAMGPDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSNS
                     TTGLIPTNTPRNSPGVPVKTGQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYYE
                     GMAKERKFNVTKEHRNQCTWGSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQA
                     SENQYLVPGYNRWWACNTGLTPCVSTSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEYD
                     YRYNRPKREPVSLTLAVMLGLGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLRA
                     LKESVSNLEESLTSLSEVVLQNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDSM
                     NKLRKKLERRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLINR
                     FVAFVRERVSAVQIMVLRQQYQGLLSQGETDL"
BASE COUNT       530 a    490 c    482 g    469 t
ORIGIN
```

```
  Query Match             24.0%;  Score 1971;  DB 14;  Length 1971;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1971;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5620 atgcatcccacgttaagccggcgccacctcccgactcggggtggagagccgaaaagactg 5679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGCATCCCACGTTAAGCCGGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGACTG 60

Qy    5680 agaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactccccag 5739
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCCCAG 120

Qy    5740 gccagtagtaaacgccttatagacagctcgaaccccccatagacctttatcccttacctgg 5799
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCCAGTAGTAAACGCCTTATAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACCTGG 180

Qy    5800 ctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcctaga 5859
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTGATTATTGACCCTGATACGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGA 240

Qy    5860 ggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaaagc 5919
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGC 300

Qy    5920 acacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaagag 5979
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGAG 360

Qy    5980 aaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacctccaac 6039
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAAC 420

Qy    6040 gatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtcaat 6099
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAAT 480

Qy    6100 tccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctccccatca 6159
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TCCGGCCCGGGCAAGTACAAAATGATGAAACTATATAAAGATAAGAGCTGCTCCCCATCA 540

Qy    6160 gacttagattatctaaagataagtttcactgaaaggaaaacaggaaaatattcaaaagtg 6219
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GACTTAGATTATCTAAAGATAAGTTTCACTGAAAGGAAAACAGGAAAATATTCAAAAGTG 600

Qy    6220 gataaatggtatgagctggggaatagttttttattatatggcgggggagcagggtccact 6279
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GATAAATGGTATGAGCTGGGGAATAGTTTTTTATTATATGGCGGGGGAGCAGGGTCCACT 660

Qy    6280 ttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatgggacccgat 6339
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAATGGGACCCGAT 720

Qy    6340 aaagtactggctgaacaggggccccggccctggagccaccgcataacttgccggtgccc 6399
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCC 780

Qy    6400 caattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccactggattg 6459
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACAGTACCACTGGATTG 840

Qy    6460 attcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacagagactc 6519
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTC 900

Qy    6520 ttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgccacttct 6579
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCT 960

Qy    6580 tcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaagaaaga 6639
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAAGA 1020

Qy    6640 aaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaataagctt 6699
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTT 1080

Qy    6700 accctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcccaccaa 6759
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAA 1140

Qy    6760 caccttgctatagtactgtggtttatgagcaggcctcagaaaatcagtatttagtacct 6819
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCT 1200

Qy    6820 ggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacctcagtc 6879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTC 1260

Qy    6880 ttcaaccaatccaaagatttctgtgtcatggtccaaatcgtcccccgagtgtactaccat 6939
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCAT 1320

Qy    6940 cctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaagagaaccc 6999
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGAACCC 1380

Qy    7000 gtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtaggaacaggg 7059
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGG 1440

Qy    7060 acagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagctacatgcg 7119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 ACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACATGCG 1500

Qy    7120 gccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagagtccctg 7179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GCCATGACAGAAGATCTCCGAGCCTTAAAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTG 1560

Qy    7180 acttctttgtctgaagtggttctacagaaccggaggggattagatctgctgtttctaaga 7239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 ACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCTAAGA 1620

Qy    7240 gaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcactcagga 7299
```

```
Db   1621 GAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTCAGGA 1680

Qy   7300 gccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagggaaaga 7359
Db   1681 GCCATCAGAGACTCCATGAACAAGCTTAGAAAAAAGTTAGAGAGGCGTCGAAGGGAAAGA 1740

Qy   7360 gaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgaccaccctg 7419
Db   1741 GAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACCACCCTG 1800

Qy   7420 ctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttgggccttgc 7479
Db   1801 CTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGGCCTTGC 1860

Qy   7480 ttaattaataggtttgttgccttttgttagagaacgagtgagtgcagtccagatcatggta 7539
Db   1861 TTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTA 1920

Qy   7540 cttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctag 7590
Db   1921 CTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAG 1971


RESULT  13
A66559
LOCUS       A66559       1974 bp    DNA             PAT       29-MAR-1999
DEFINITION  Sequence 9 from Patent WO9740167.
ACCESSION   A66559
VERSION     A66559.1  GI:4538112
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1974)
  AUTHORS   Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
  TITLE     PORCINE RETROVIRUS
  JOURNAL   Patent: WO 9740167-A 9 30-OCT-1997;
            Q ONE BIOTECH LTD (GB)
FEATURES             Location/Qualifiers
     source          1. .1974
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      534 a    489 c    483 g    468 t
ORIGIN


  Query Match           18.3%;  Score 1499;  DB 6;  Length 1974;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1949;  Conservative    0;  Mismatches    3;  Indels    3;  Gaps    2;

Qy   5639 ggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcg 5698
Db     20 GGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGACTGAGAATCCCCTTAAGCTTCG 79

Qy   5699 cctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttta 5758
Db     80 CCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTA 139

Qy   5759 tagacagctcgaaccccatagaccttatcccttacctggctgattattgaccctgata 5818
Db    140 TAGACAGCTCGAACCCCCATAGACCTTTATCCCCTACCTGGCTGATTATTGACCCTGATA 199

Qy   5819 cgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac 5878
Db    200 CGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGTGGCCTGAAC 259

Qy   5879 tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtcc 5938
Db    260 TGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGCACACCTCCCAACCTAGTCC 319

Qy   5939 gtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctg 5998
Db    320 GTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGAGAAATACTGTGGGGGTTCTG 379

Qy   5999 gggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggc 6058
Db    380 GGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGC 439

Qy   6059 cgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtaca 6118
Db    440 CGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACA 499

Qy   6119 aaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaaga 6178
Db    500 AAATGATGAAACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTATCTAAAGA 559

Qy   6179 taagtttcactg--aaaggaaaacaggaaaatattcaaaagtggataaatggtatgagct 6236
Db    560 TAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTATGAGCT 619

Qy   6237 ggggaatagttttttat-tatatggcgggggagcagggtccactttaaccattcgccttta 6295
Db    620 GGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTCCACTTTAACCATTCGCCTTA 679

Qy   6296 ggatagagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaac 6355
Db    680 GGATAGAGACGGGGACAGAACCCCCTGTGGCAATGGGACCCGATAAAGTACTGGCTGAAC 739

Qy   6356 aggggcccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgc 6415
Db    740 AGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGC 799

Qy   6416 ggcctgacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgc 6475
Db    800 GGCCTGACATAACACAGCCGCCTAGCAACAGTACCACTGGATTGATTCCTACCAACACGC 859

Qy   6476 ctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagg 6535
Db    860 CTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGG 919

Qy   6536 gagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtc 6595
Db    920 GAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTTGTC 979

Qy   6596 tatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgacca 6655
Db    980 TATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAAGAAAATTCAATGTGACCA 1039

Qy   6656 aagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagttt 6715
Db   1040 AAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTT 1099

Qy   6716 ccgggaaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagta 6775
Db   1100 CCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAGTA 1159

Qy   6776 ctgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggt 6835
Db   1160 CTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGT 1219

Qy   6836 gggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaag 6895
Db   1220 GGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAG 1279

Qy   6896 atttctgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtcc 6955
Db   1280 ATTTGTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCC 1339

Qy   6956 ttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctag 7015
Db   1340 TTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAG 1399

Qy   7016 ctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatca 7075
Db   1400 CTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTGCCCTGATCA 1459
```

```
Qy   7076 caggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatc 7135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1460 CAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGATC 1519

Qy   7136 tccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaag 7195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1520 TCCGAGCCTTAAAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAG 1579

Qy   7196 tggttctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtg 7255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1580 TGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTG 1639

Qy   7256 cagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactcca 7315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1640 CAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCA 1699

Qy   7316 tgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggt 7375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1700 TGAACAAGCTTAGAAAAAAGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGT 1759

Qy   7376 ggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacgg 7435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1760 GGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGG 1819

Qy   7436 ggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttg 7495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1820 GGCCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTG 1879

Qy   7496 ttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtacc 7555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1880 TTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACC 1939

Qy   7556 aaggccttctgagccaaggagaaactgacctctag 7590
          ||||||||||||||||||||||||||||||||||
Db   1940 AAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAG 1974


RESULT  14
AX002804
LOCUS       AX002804     3482 bp    DNA             PAT       21-AUG-2000
DEFINITION  Sequence 3 from Patent WO9853104.
ACCESSION   AX002804
VERSION     AX002804.1  GI:9885132
KEYWORDS    .
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 3482)
  AUTHORS   Stoye,J.P. and Weiss,R.A.
  TITLE     Detection of retroviral subtypes based upon envelope specific
            sequences
  JOURNAL   Patent: WO 9853104-A 3 26-NOV-1998;
            MEDICAL RES COUNCIL (GB); STOYE JONATHAN PAUL (GB)
FEATURES             Location/Qualifiers
     source          1. .3482
                     /organism="porcine endogenous retrovirus"
                     /db_xref="taxon:61673"
BASE COUNT      927 a    854 c    867 g    834 t
ORIGIN


  Query Match             17.9%;  Score 1466;  DB 6;  Length 3482;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 2536;  Conservative    0;  Mismatches   13;  Indels    4;  Gaps    3;

Qy   5639 ggcgccacctcccgactcggggtggagagccgaaaagactgagaatcccccttaagcttcg 5698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    930 GGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGACTGAGAATCCCCTTAAGCTTCG 989

Qy   5699 cctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttа 5758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    990 CCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTA 1049

Qy   5759 tagacagctcgaaccccccatagacctttatcccttacctggctgattattgaccctgata 5818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1050 TAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGATA 1109

Qy   5819 cgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac 5878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1110 CGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGTGGCCTGAAC 1169

Qy   5879 tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtcc 5938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 TGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGCACACCTCCCAACCTAGTCC 1229

Qy   5939 gtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctg 5998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 GTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGAGAAATACTGTGGGGGTTCTG 1289

Qy   5999 gggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggc 6058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 GGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGC 1349

Qy   6059 cgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtaca 6118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1350 CGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACA 1409

Qy   6119 aaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaaga 6178
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1410 AAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTATCTAAAGA 1469

Qy   6179 taagtttcactg--aaaggaaaacaggaaaatattcaaaagtggataaatggtatgagct 6236
          |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1470 TAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTATGAGCT 1529

Qy   6237 ggggaatagttttttat-tatatggcgggggagcagggtccactttaaccattcgccttа 6295
          |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db   1530 GGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTCCACTTTAACCATTCGCCTTA 1589

Qy   6296 ggatagagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaac 6355
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1590 GGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACCCGATAAAGTACTGGCTGAAC 1649

Qy   6356 aggggccccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgc 6415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1650 AGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGC 1709

Qy   6416 ggcctgacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgc 6475
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1710 GGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGGATTGATTCCTACCAACACGC 1769

Qy   6476 ctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagg 6535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1770 CTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGG 1829

Qy   6536 gagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtc 6595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1830 GAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTTGTC 1889

Qy   6596 tatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgacca 6655
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1890 TATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAGGAAAATTCAATGTGACCA 1949

Qy   6656 aagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagttt 6715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1950 AAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTT 2009

Qy   6716 ccgggaaggggacatgcataggaaaagctccccccatcccaccaacacctttgctatagta 6775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2010 CCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAGTA 2069

Qy   6776 ctgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggt 6835
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2070 CTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGT 2129
```

```
Qy   6836 gggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaag 6895
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2130 GGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAG 2189

Qy   6896 atttctgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtcc 6955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2190 ATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCC 2249

Qy   6956 ttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctag 7015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2250 TTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAG 2309

Qy   7016 ctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatca 7075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2310 CTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTGCCCTGATCA 2369

Qy   7076 caggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatc 7135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2370 CAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGATC 2429

Qy   7136 tccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaag 7195
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2430 TCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAG 2489

Qy   7196 tggttctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtg 7255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2490 TGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTG 2549

Qy   7256 cagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactcca 7315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2550 CAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCA 2609

Qy   7316 tgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggt 7375
          ||| |||||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2610 TGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGT 2669

Qy   7376 ggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacgg 7435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2670 GGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGG 2729

Qy   7436 ggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttg 7495
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2730 GACCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTG 2789

Qy   7496 ttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtacc 7555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2790 TTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACC 2849

Qy   7556 aaggccttctgagccaaggagaaactgacctctagccttcccagttctaagattagaact 7615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2850 AAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACT 2909

Qy   7616 attaacaagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaac 7675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2910 ATTAACAAGACAAGAAGTGGGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAAC 2969

Qy   7676 ccaggaagttaataaaaagctctaaatgcccccgaattacagaccctgctggctgccagt 7735
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   2970 CCAGGAAGTTAATAAAAAGCTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGT 3029

Qy   7736 aaataggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaag 7795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3030 AAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAG 3089

Qy   7796 ataacaggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccat 7855
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   3090 ATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAAAACCGACTGGCACCAT 3149

Qy   7856 agaagaattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactc 7915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3150 AGAAGAATTGATTACACATTGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTC 3209
```

```
Qy   7916 tgcccaggagcccacgcagatgcggacctccggagctattttaaaatgattggtccacgg 7975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3210 TGCCCAGGAGCCCACGCAGATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGG 3269

Qy   7976 agcgcgggctctcgatattttaaaatgattggtccatggagcgcgggctc-tcgatattt 8034
          |||||||||||||||||||||||||||||||||||| |||||||||||||| |||||||||
Db   3270 AGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTTCGATATTT 3329

Qy   8035 taaaatgattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgat 8094
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3330 TAAAATGATTGGTTTGTGACGCACAGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGAT 3389

Qy   8095 tccgcactcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgc 8154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3390 TCCGCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGC 3449

Qy   8155 gcttggaataaaaatcctcttgctgtttgcatc 8187
          |||||||||||||||||||||||||||||||
Db   3450 GCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 3482
```

```
RESULT  15
PERENV2
LOCUS       PERENV2      3482 bp    RNA             VRL       14-OCT-1997
DEFINITION  Porcine endogenous retrovirus env gene, 3482 bp.
ACCESSION   Y12239
VERSION     Y12239.1  GI:2576324
KEYWORDS    env gene.
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 3482)
  AUTHORS   Le Tissier,P., Stoye,J.P., Takeuchi,Y., Patience,C. and Weiss,R.A.
  TITLE     Two sets of human-tropic pig retrovirus
  JOURNAL   Nature 389 (6652), 681-682 (1997)
  MEDLINE   97478526
REFERENCE   2  (bases 1 to 3482)
  AUTHORS   Stoye,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-1997) J.P. Stoye, National Institute for Medical
            Research, Virology, The Ridgeway Mill Hill, London, NW7 1AA, UK
  REMARK    revised by author 16-SEP-97
FEATURES             Location/Qualifiers
     source          1. .3482
                     /organism="porcine endogenous retrovirus"
                     /specific_host="Sus scrofa"
                     /db_xref="taxon:61673"
                     /note="PERV-B"
     gene            911. .2884
                     /gene="env"
     CDS             911. .2884
                     /gene="env"
                     /codon_start=1
                     /protein_id="CAA72928.1"
                     /db_xref="GI:2576325"
                     /db_xref="SPTREMBL:O41173"
                     /translation="MHPTLSWRHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSK
                     RLIDSSNPHRPLSLTWLIIDPDTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTP
                     PNLVRSYGFYCCPGTEKEKYCGGSGESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVN
                     SGPGKYKVMKLYKDKSCSPSDLDYLKISFTEKGKQENIQKWINGMSWGIVFYKYGGGA
                     GSTLTIRLRIETGTEPPVAVGPDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSN
                     GTTGLIPTNTPRNSPGVPVKTGQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYY
                     EGMAKEGKFNVTKEHRNQCTWGSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQ
                     ASENQYLVPGYNRWWACNTGLTPCVSTSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEY
                     DYRYNRPKREPVSLTLAVMLGLGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLR
                     ALEESVSNLEESLTSLSEVVLQNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDS
                     MSKLRERLERRRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLIN
                     RFVAFVRERVSAVQIMVLRQQYQGLLSQGETDL"
BASE COUNT      927 a    854 c    867 g    834 t
ORIGIN
```

Query Match 17.9%; Score 1466; DB 14; Length 3482;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2536; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

```
Qy   5639 ggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcg 5698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    930 GGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGACTGAGAATCCCCTTAAGCTTCG 989

Qy   5699 cctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttta 5758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    990 CCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTA 1049

Qy   5759 tagacagctcgaaccccccatagaccttttatcccttacctggctgattattgaccctgata 5818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1050 TAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGATA 1109

Qy   5819 cgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac 5878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1110 CGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGTGGCCTGAAC 1169

Qy   5879 tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtcc 5938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 TGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGCACACCTCCCAACCTAGTCC 1229

Qy   5939 gtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctg 5998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 GTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGAGAAATACTGTGGGGGTTCTG 1289

Qy   5999 gggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggc 6058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 GGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGC 1349

Qy   6059 cgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtaca 6118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1350 CGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACA 1409

Qy   6119 aaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaaga 6178
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1410 AAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTATCTAAAGA 1469

Qy   6179 taagtttcactg--aaaggaaaacaggaaaatattcaaaagtggataaatggtatgagct 6236
          |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db   1470 TAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTATGAGCT 1529

Qy   6237 ggggaatagttttttat-tatatggcgggggagcagggtccacttttaaccattcgccttta 6295
          ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db   1530 GGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTCCACTTTAACCATTCGCCTTA 1589

Qy   6296 ggatagagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaac 6355
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1590 GGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGACCCGATAAAGTACTGGCTGAAC 1649

Qy   6356 aggggcccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgc 6415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1650 AGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGC 1709

Qy   6416 ggcctgacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgc 6475
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1710 GGCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGGATTGATTCCTACCAACACGC 1769

Qy   6476 ctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagg 6535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1770 CTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGG 1829

Qy   6536 gagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtc 6595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1830 GAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTTGTC 1889

Qy   6596 tatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgacca 6655
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1890 TATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAGGAAAATTCAATGTGACCA 1949

Qy   6656 aagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagttt 6715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1950 AAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTT 2009

Qy   6716 ccgggaaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagta 6775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2010 CCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAGTA 2069

Qy   6776 ctgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggt 6835
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2070 CTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGT 2129

Qy   6836 gggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaag 6895
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2130 GGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAG 2189

Qy   6896 atttctgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtcc 6955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2190 ATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCC 2249

Qy   6956 ttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctag 7015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2250 TTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAG 2309

Qy   7016 ctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatca 7075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2310 CTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTGCCCTGATCA 2369

Qy   7076 caggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatc 7135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2370 CAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGATC 2429

Qy   7136 tccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaag 7195
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   2430 TCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAG 2489

Qy   7196 tggttctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtg 7255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2490 TGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTG 2549

Qy   7256 cagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactcca 7315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2550 CAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCA 2609

Qy   7316 tgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggt 7375
          ||| |||||||||| ||| ||||||||||||||||||||||||||||||||||||||||
Db   2610 TGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGT 2669

Qy   7376 ggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacgg 7435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2670 GGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGG 2729

Qy   7436 ggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttg 7495
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2730 GACCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTG 2789

Qy   7496 ttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtacc 7555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2790 TTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACC 2849

Qy   7556 aaggccttctgagccaaggagaaactgacctctagccttcccagttctaagattagaact 7615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2850 AAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACT 2909

Qy   7616 attaacaagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaac 7675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2910 ATTAACAAGACAAGAAGTGGGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAAC 2969

Qy   7676 ccaggaagttaataaaaagctctaaatgcccccgaattacagaccctgctggctgccagt 7735
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   2970 CCAGGAAGTTAATAAAAAGCTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGT 3029

Qy   7736 aaataggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaag 7795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    3030 AAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAG 3089

Qy    7796 ataacaggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccat 7855
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    3090 ATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAAAACCGACTGGCACCAT 3149

Qy    7856 agaagaattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactc 7915
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3150 AGAAGAATTGATTACACATTGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTC 3209

Qy    7916 tgcccaggagcccacgcagatgcggacctccggagctattttaaaatgattggtccacgg 7975
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3210 TGCCCAGGAGCCCACGCAGATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGG 3269

Qy    7976 agcgcgggctctcgatattttaaaatgattggtccatggagcgcgggctc-tcgatattt 8034
           |||||||||||||||||||||||||||||||||| |||||||||||||| |||||||||
Db    3270 AGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTTCGATATTT 3329

Qy    8035 taaaatgattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgat 8094
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3330 TAAAATGATTGGTTTGTGACGCACAGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGAT 3389

Qy    8095 tccgcactcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgc 8154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3390 TCCGCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGC 3449

Qy    8155 gcttggaataaaaatcctcttgctgtttgcatc 8187
           |||||||||||||||||||||||||||||||
Db    3450 GCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 3482
```

Search completed: February 24, 2002, 07:30:19
Job time: 28334 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:30:19 ; Search time 10173.5 Seconds
(without alignments)
4793.420 Million cell updates/sec

Title: US-09-171-553B-9
Perfect score: 2956
Sequence: 1 tgctttttagggttaggaac.....aaaaaaaaaaaaaaaaaaaa 2956

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.	Score	Query Match	Length	DB	ID	Description
1	2191	74.1	8209	6	A66553	A66553 Sequence 3
2	2089	70.7	8196	6	A66552	A66552 Sequence 2
3	1872	63.3	1974	6	A66559	A66559 Sequence 9
4	1763	59.6	8750	14	PEN133818	AJ133818 Porcine e
5	1743	59.0	3482	6	AX002804	AX002804 Sequence
6	1743	59.0	3482	14	PERENV2	Y12239 Porcine end
7	1654	56.0	7808	14	PERY17013	Y17013 Porcine end
8	1636	55.3	8918	14	PEN133816	AJ133816 Porcine e
9	1572	53.2	8763	14	PEN293657	AJ293657 Porcine e
10	1499	50.7	1971	14	AF014162	AF014162 Porcine e
11	813	27.5	7333	4	AF038601	AF038601 Sus scrof
12	813	27.5	7333	6	AR130474	AR130474 Sequence
13	626	21.2	8849	14	PEN133817	AJ133817 Porcine e
14	457	15.5	2462	6	AX002802	AX002802 Sequence
15	457	15.5	2462	14	PERENV1	Y12238 Porcine end
16	446	15.1	1042	14	PERY17012	Y17012 Porcine end
17	412	13.9	2000	14	PEN288588	AJ288588 Porcine e
18	375	12.7	3025	4	AF147808	AF147808 Sus scrof
19	347	11.7	1968	14	PEN288589	AJ288589 Porcine e
20	321	10.9	746	14	PEN298073	AJ298073 Porcine e
21	318	10.8	590	14	PEN298074	AJ298074 Porcine e
22	318	10.8	2000	14	PEN288592	AJ288592 Porcine e
23	143	4.8	3320	6	A66551	A66551 Sequence 1
24	143	4.8	8918	14	PEN293656	AJ293656 Porcine e
25	139	4.7	2002	14	PEN288591	AJ288591 Porcine e
26	129	4.4	1996	14	PEN288590	AJ288590 Porcine e
27	116	3.9	1999	14	PEN288587	AJ288587 Porcine e
28	98	3.3	8132	4	AF038600	AF038600 Sus scrof
29	98	3.3	8132	6	AR130475	AR130475 Sequence
30	91	3.1	633	6	AX052622	AX052622 Sequence
31	91	3.1	704	6	AX052621	AX052621 Sequence
32	91	3.1	4918	6	AX052637	AX052637 Sequence
33	91	3.1	6076	6	AX052636	AX052636 Sequence
34	91	3.1	6076	6	AX052647	AX052647 Sequence
35	91	3.1	7362	6	AX052634	AX052634 Sequence
36	91	3.1	7873	6	AX052638	AX052638 Sequence
37	91	3.1	8060	6	AR130473	AR130473 Sequence
38	91	3.1	8132	4	AF038599	AF038599 Sus scrof
39	88	3.0	2008	14	PEN288586	AJ288586 Porcine e
40	74	2.5	4402	6	AX052635	AX052635 Sequence
41	62	2.1	1996	14	PEN288584	AJ288584 Porcine e
42	57	1.9	610	4	AF286463	AF286463 Sus scrof
43	53	1.8	595	14	PEN298075	AJ298075 Porcine e
44	41	1.4	1591	10	AF161307	AF161307 Mus carol
45	41	1.4	195423	2	AC087151	AC087151 Mus muscu

ALIGNMENTS

RESULT 1
A66553
LOCUS A66553 8209 bp DNA PAT 29-MAR-1999
DEFINITION Sequence 3 from Patent WO9740167.
ACCESSION A66553
VERSION A66553.1 GI:4538106
KEYWORDS .
SOURCE porcine endogenous retrovirus.
ORGANISM porcine endogenous retrovirus
Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses; 1-Mammalian type C virus group.
REFERENCE 1 (bases 1 to 8209)
AUTHORS Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
TITLE PORCINE RETROVIRUS
JOURNAL Patent: WO 9740167-A 3 30-OCT-1997;
Q ONE BIOTECH LTD (GB)
FEATURES Location/Qualifiers
source 1. .8209
/organism="porcine endogenous retrovirus"
/db_xref="taxon:61673"

```
     CDS             588. .2162
                     /codon_start=1
                     /product="VIRAL CORE PROTEIN"
                     /protein_id="CAB39347.1"
                     /db_xref="GI:4538107"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPGSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSSSYLPRDRGAADLAGTPTCSPTPLSSTGCCEG
                     TSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREEEREERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKD"
     CDS             <2163. .5747
                     /codon_start=1
                     /product="REVERSE TRANSCRIPTASE"
                     /protein_id="CAB39348.1"
                     /db_xref="GI:4538108"
                     /translation="GRRGSDPLPEPRVTLKVEGQPVEFLVDTGAEHSVLLQPLGKLKE
                     KKSWVMGATGQRQYPWTTRRTVDLGVGRVTHSFLVIPECPVPLLGRDLLTKMGAQISF
                     EQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLA
                     KQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLP
                     VRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCL
                     RLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQ
                     VTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLR
                     GGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKG
                     GFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRR
                     PVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPP
                     DRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVR
                     KDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELM
                     ALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLE
                     ALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPIIETPKAPEPRRQ
                     YTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEGLEYVQQIHRLTHLGTKHLQQL
                     VRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRIPPGKRLRGSHPGAHWEVDFTEVK
                     PAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKILEEIFPRFGIPKVIGSDNGPA
                     FVAQVSQGLAKILGIDWKLHCAYRPQSSGQVERMNRTIKETLTKLTTETGINDWMALL
                     PFVLFRVRNTPGQPGLTPYELLYGGPPPLAEIAFAHSADVLLSQPLFSRLKALEWVRQ
                     RAWKQLREAYSGGDLQVPHRFQVGDSVYVRRHRAGNLETRWKGPYLVLLTTPTAVKVE
                     GIPTWIHASHVKPAPPPDSGWRAEKTENPLKLRLHRLVPYSNNNSPGQ"
     CDS             5620. .7590
                     /codon_start=1
                     /product="VIRAL ENVELOPE PROTEIN"
                     /protein_id="CAB39349.1"
                     /db_xref="GI:4538109"
                     /translation="MHPTLSRRHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSK
                     RLIDSSNPHRPLSLTWLIIDPDTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTP
                     PNLVRSYGFYCCPGTEKEKYCGGSGESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVN
                     SGPGKYKMMKLYKDKSCSPSDLDYLKISFTERKTGKYSKVDKWYELGNSFLLYGGGAG
                     STLTIRLRIETGTEPPVAMGPDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSNS
                     TTGLIPTNTPRNSPGVPVKTGQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYYE
                     GMAKERKFNVTKEHRNQCTWGSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQA
                     SENQYLVPGYNRWWACNTGLTPCVSTSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEYD
                     YRYNRPKREPVSLTLAVMLGLGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLRA
                     LKESVSNLEESLTSLSEVVLQNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDSM
                     NKLRKKLERRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLINR
                     FVAFVRERVSAVQIMVLRQQYQGLLSQGETDL"
BASE COUNT     2168 a   2064 c   2154 g   1823 t
ORIGIN
```

Query Match 74.1%; Score 2191; DB 6; Length 8209;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2741; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

```
Qy    208 gagacttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgccacc 267
Db   5464 GAGACTTGCAAGTTCCACATCGCTTCCAAGTTGGAGATTCAGTCTATGTTAGACGCCACC 5523

Qy    268 gtgcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaa 327
Db   5524 GTGCAGGAAACCTCGAGACTCGGTGGAAGGGACCTTATCTCGTACTTTTGACCACACCAA 5583

Qy    328 cggctgtgaaagtcgaaggaatctccacctggatccatgcatcccacgttaagctggcgc 387
Db   5584 CGGCTGTGAAAGTCGAAGGAATCCCCACCTGGATCCATGCATCCCACGTTAAGCCGGCGC 5643

Qy    388 cacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcgcctcc 447
Db   5644 CACCTCCCGACTCGGGGTGGAGAGCCGAAAAGACTGAGAATCCCCTTAAGCTTCGCCTCC 5703

Qy    448 atcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttatagac 507
Db   5704 ATCGCCTGGTTCCTTACTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTATAGAC 5763

Qy    508 agctcgaaccccccatagacctttatcccctacctggctgattattgaccctgatacgggt 567
Db   5764 AGCTCGAACCCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGATACGGGT 5823

Qy    568 gtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaactgcat 627
Db   5824 GTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGTGGCCTGAACTGCAT 5883

Qy    628 ttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtccgtagt 687
Db   5884 TTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGCACACCTCCCAACCTAGTCCGTAGT 5943

Qy    688 tatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctgggaa 747
Db   5944 TATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGAGAAATACTGTGGGGGTTCTGGGGAA 6003

Qy    748 tccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggccgatc 807
Db   6004 TCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGCCGATC 6063

Qy    808 tctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtacaaaatg 867
Db   6064 TCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACAAAATG 6123

Qy    868 atgaaactatataaagataagagctgctccccatcagacttagattatctaaagataagt 927
Db   6124 ATGAAACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTATCTAAAGATAAGT 6183

Qy    928 ttcactgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagctgggga 987
Db   6184 TTCACTG--AAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTATGAGCTGGGGA 6241

Qy    988 atagttttttaataatatggcgggggagcagggtccacttaaccattcgccttaggata 1047
Db   6242 ATAGTTTTTT-ATTATATGGCGGGGGAGCAGGGTCCACTTTAACCATTCGCCTTAGGATA 6300

Qy   1048 gagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaacagggg 1107
Db   6301 GAGACGGGGACAGAACCCCCTGTGGCAATGGGACCCGATAAAGTACTGGCTGAACAGGGG 6360

Qy   1108 cccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgcggcct 1167
Db   6361 CCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGCGGCCT 6420

Qy   1168 gacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgcctaga 1227
Db   6421 GACATAACACAGCCGCCTAGCAACAGTACCACTGGATTGATTCCTACCAACACGCCTAGA 6480

Qy   1228 aactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagggagct 1287
Db   6481 AACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGGGAGCT 6540

Qy   1288 ttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtctatcc 1347
Db   6541 TTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTTGTCTATCC 6600

Qy   1348 tcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagag 1407
Db   6601 TCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAAGAAAATTCAATGTGACCAAAGAG 6660
```

```
Qy   1408 catagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagtttccggg 1467
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6661 CATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTTCCGGG 6720

Qy   1468 aaggggacatgcataggaaaagctccccccatcccaccaacaccttttgctatagtactgtg 1527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6721 AAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAGTACTGTG 6780

Qy   1528 gtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggca 1587
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6781 GTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGTGGGCA 6840

Qy   1588 tgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttg 1647
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6841 TGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAGATTTC 6900

Qy   1648 tgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgat 1707
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6901 TGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCCTTGAT 6960

Qy   1708 gaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctagctgta 1767
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6961 GAATATGACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAGCTGTA 7020

Qy   1768 atgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatcacagga 1827
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7021 ATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTGCCCTGATCACAGGA 7080

Qy   1828 ccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatctccga 1887
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7081 CCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGATCTCCGA 7140

Qy   1888 gccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaagtggtt 1947
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7141 GCCTTAAAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAGTGGTT 7200

Qy   1948 ctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtgcagcc 2007
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7201 CTACAGAACCGGAGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTGCAGCC 7260

Qy   2008 ttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactccatgaac 2067
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7261 TTAAAAGAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCATGAAC 7320

Qy   2068 aagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggtggttt 2127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7321 AAGCTTAGAAAAAAGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTT 7380

Qy   2128 gaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggccc 2187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7381 GAAGGATGGTTCAACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCC 7440

Qy   2188 ctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgcc 2247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7441 CTAGTAGTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCC 7500

Qy   2248 tttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggc 2307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7501 TTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGC 7560

Qy   2308 cttctgagccaaggagaaactgacctctagccttcccagttctaagattagaactattaa 2367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7561 CTTCTGAGCCAAGGAGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAA 7620

Qy   2368 caagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccagg 2427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7621 CAAGACAAGAAGTGGGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGG 7680

Qy   2428 aagttaataaaaagctctaaatgcccccgaattacagaccctgctggctgccagtaaata 2487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7681 AAGTTAATAAAAAGCTCTAAATGCCCCCGAATTACAGACCCTGCTGGCTGCCAGTAAATA 7740

Qy   2488 ggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaagataac 2547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7741 GGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAAC 7800

Qy   2548 aggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccatagaag 2607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7801 AGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACCATAGAAG 7860

Qy   2608 aattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactctgccc 2667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7861 AATTGATTACACATTGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTCTGCCC 7920

Qy   2668 aggagcccacgcagatgcggacctccggagctattttaaaatgattggtccacggagcgc 2727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7921 AGGAGCCCACGCAGATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGGAGCGC 7980

Qy   2728 gggctctcgatattttaaaatgattggtccatggagcgcgggctctcgatattttaaaat 2787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7981 GGGCTCTCGATATTTTAAAATGATTGGTCCATGGAGCGCGGGCTCTCGATATTTTAAAAT 8040

Qy   2788 gattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgattccgca 2847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8041 GATTGGTTTGTGACGCACAGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCGCA 8100

Qy   2848 ctcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgcttgg 2907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8101 CTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCTTGG 8160

Qy   2908 aataaaaatcctcttgctgtttgcatcaaaaaaaaaaaaaaaaaaaaaa 2956
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   8161 AATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAAAAAAAAAAAAAAA 8209
```

RESULT 2
A66552

```
LOCUS       A66552       8196 bp    DNA             PAT       29-MAR-1999
DEFINITION  Sequence 2 from Patent WO9740167.
ACCESSION   A66552
VERSION     A66552.1  GI:4538105
KEYWORDS    .
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 8196)
  AUTHORS   Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
  TITLE     PORCINE RETROVIRUS
  JOURNAL   Patent: WO 9740167-A 2 30-OCT-1997;
            Q ONE BIOTECH LTD (GB)
FEATURES             Location/Qualifiers
     source          1. .8196
                     /organism="porcine endogenous retrovirus"
                     /db_xref="taxon:61673"
BASE COUNT     2165 a    2061 c    2146 g    1820 t      4 others
ORIGIN
```

Query Match 70.7%; Score 2089; DB 6; Length 8196;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2739; Conservative 0; Mismatches 7; Indels 3; Gaps 2;

```
Qy    208 gagacttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgccacc 267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5450 GAGACTTGCAAGTTCCACATCGCTTCCAAGTTGGAGATTCAGTCTATGTTAGACGCCACC 5509

Qy    268 gtgcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaa 327
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5510 GTGCAGGAAACCTCGAGACTCGGTGGAAGGGACCTTATCTCGTACTTTTGACCACACCAA 5569

Qy    328 cggctgtgaaagtcgaaggaatctccacctggatccatgcatcccacgttaagctggcgc 387
          |||||||||||||||||||||| |||||||||||||||||||||||||||||| |||||
Db   5570 CGGCTGTGAAAGTCGAAGGAATCCCCACCTGGATCCATGCATCCCACGTTAAGCYGGCGC 5629

Qy    388 cacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcgcctcc 447
```

```
Db  5630 CACCTCCCGACTCGGGGTGGAGAGCCGAAAAGACTGAGAATCCCCTTAAGCTTCGCCTCC 5689

Qy   448 atcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttatagac 507
Db  5690 ATCGCCTGGTTCCTTACTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTATAGAC 5749

Qy   508 agctcgaacccccatagacctttatcccctacctggctgattattgaccctgatacgggt 567
Db  5750 AGCTCGAACCCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGATACGGGT 5809

Qy   568 gtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaactgcat 627
Db  5810 GTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGTGGCCTGAACTGCAT 5869

Qy   628 ttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtccgtagt 687
Db  5870 TTCTGCCTCCGATTGATTAACCCCGCTGTTAARAGCACACCTCCCAACCTAGTCCGTAGT 5929

Qy   688 tatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctggggaa 747
Db  5930 TATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGAGAAATACTGTGGGGGTTCTGGGGAA 5989

Qy   748 tccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggccgatc 807
Db  5990 TCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGCCGATC 6049

Qy   808 tctctccaggaccgggtaaaattctccttttgtcaattccggcccgggcaagtacaaaatg 867
Db  6050 TCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACAAAATG 6109

Qy   868 atgaaactatataaagataagagctgctccccatcagacttagattatctaaagataagt 927
Db  6110 ATGAAACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTATCTAAAGATAAGT 6169

Qy   928 ttcactgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagctgggga 987
Db  6170 TTCACTG--AAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTATGAGCTGGGGA 6227

Qy   988 atagtttttaataatatggcgggggagcagggtccacttaaccattcgccttaggata 1047
Db  6228 ATAGTTTTTT-ATTATATGGCGGGGGAGCAGGGTCCACTTTAACCATTCGCCTTAGGATA 6286

Qy  1048 gagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaacagggg 1107
Db  6287 GAGACGGGGACAGAACCCCCTGTGGCAATGGGACCCGATAAAGTACTGGCTGAACAGGGG 6346

Qy  1108 cccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgcggcct 1167
Db  6347 CCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGCGGCCT 6406

Qy  1168 gacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgcctaga 1227
Db  6407 GACATAACACAGCCGCCTAGCAACAGTACCACTGGATTGATTCCTACCAACACGCCTAGA 6466

Qy  1228 aactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagggagct 1287
Db  6467 AACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGGGAGCT 6526

Qy  1288 ttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtctatcc 1347
Db  6527 TTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTTGTCTATCC 6586

Qy  1348 tcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagag 1407
Db  6587 TCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAAGAAAATTCAATGTGACCAAAGAG 6646

Qy  1408 catagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagtttccggg 1467
Db  6647 CATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTTCCGGG 6706

Qy  1468 aaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagtactgtg 1527
Db  6707 AAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAGTACTGTG 6766

Qy  1528 gtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggca 1587
Db  6767 GTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGTGGGCA 6826

Qy  1588 tgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttg 1647
Db  6827 TGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAGATTTC 6886

Qy  1648 tgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgat 1707
Db  6887 TGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCCTTGAT 6946

Qy  1708 gaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctagctgta 1767
Db  6947 GAATATGACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAGCTGTA 7006

Qy  1768 atgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatcacagga 1827
Db  7007 ATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTGCCCTGATCACAGGA 7066

Qy  1828 ccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatctccga 1887
Db  7067 CCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGATCTCCGA 7126

Qy  1888 gccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaagtggtt 1947
Db  7127 GCCTTAAAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAGTGGTT 7186

Qy  1948 ctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtgcagcc 2007
Db  7187 CTACAGAACCGGAGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTGCAGCC 7246

Qy  2008 ttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactccatgaac 2067
Db  7247 TTAAAAGAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCATGAAC 7306

Qy  2068 aagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggtggttt 2127
Db  7307 AAGCTTAGAAAAAAGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTT 7366

Qy  2128 gaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggccc 2187
Db  7367 GAAGGATGGTTCAACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCC 7426

Qy  2188 ctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgcc 2247
Db  7427 CTAGTAGTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCC 7486

Qy  2248 tttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggc 2307
Db  7487 TTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGC 7546

Qy  2308 cttctgagccaaggagaaactgacctctagccttcccagttctaagattagaactattaa 2367
Db  7547 CTTCTGAGCCAAGGAGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAA 7606

Qy  2368 caagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccagg 2427
Db  7607 CAAGACAAGAAGTGGGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGG 7666

Qy  2428 aagttaataaaaagctctaaatgcccccgaattacagaccctgctggctgccagtaaata 2487
Db  7667 AAGTTAATAAAAAGCTCTAAATGCCCCCGAATTMCAGACCCTGCTGGCTGCCAGTAAATA 7726

Qy  2488 ggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaagataac 2547
Db  7727 GGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAAC 7786

Qy  2548 aggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccatagaag 2607
Db  7787 AGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACCATAGAAG 7846
```

```
Qy   2608 aattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactctgccc 2667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7847 AATTGATTACACATTGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTCTGCCC 7906

Qy   2668 aggagcccacgcagatgcggacctccggagctattttaaaatgattggtccacggagcgc 2727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7907 AGGAGCCCACGCAGATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGGAGCGC 7966

Qy   2728 gggctctcgatattttaaaatgattggtccatggagcgcgggctctcgatattttaaaat 2787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7967 GGGCTCTCGATATTTTAAAATGATTGGTCCATGGAGCGCGGGCTCTCGATATTTTAAAAT 8026

Qy   2788 gattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgattccgca 2847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8027 GATTGGTTTGTGACGCACAGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCGCA 8086

Qy   2848 ctcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgcttgg 2907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8087 CTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCTTGG 8146

Qy   2908 aataaaaatcctcttgctgtttgcatcaaaaaaaaaaaaaaaaaaaaa 2956
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   8147 AATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAAAAAAAAAAAAAA 8195


RESULT   3
A66559
LOCUS       A66559       1974 bp    DNA             PAT       29-MAR-1999
DEFINITION  Sequence 9 from Patent WO9740167.
ACCESSION   A66559
VERSION     A66559.1  GI:4538112
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1974)
  AUTHORS   Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
  TITLE     PORCINE RETROVIRUS
  JOURNAL   Patent: WO 9740167-A 9 30-OCT-1997;
            Q ONE BIOTECH LTD (GB)
FEATURES             Location/Qualifiers
     source          1. .1974
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      534 a    489 c    483 g    468 t
ORIGIN


  Query Match           63.3%;  Score 1872;  DB 6;  Length 1974;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1972;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    364 atgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaagactg 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGACTG 60

Qy    424 agaatcccctaagcttcgcctccatcgcctggttccttactctaacaataactccccag 483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCCCAG 120

Qy    484 gccagtagtaaacgccttatagacagctcgaaccccccatagacctttatcccctacctgg 543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCCAGTAGTAAACGCCTTATAGACAGCTCGAACCCCCATAGACCTTTATCCCCTACCTGG 180

Qy    544 ctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcctaga 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGATTATTGACCCTGATACGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGA 240

Qy    604 ggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaaagc 663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGC 300

Qy    664 acacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaagag 723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGAG 360

Qy    724 aaatactgtgggggttctgggaatccttctgtaggagatggagctgcgtcacctccaac 783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAAC 420

Qy    784 gatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtcaat 843
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAAT 480

Qy    844 tccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctccccatca 903
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCCGGCCCGGGCAAGTACAAAATGATGAAACTATATAAAGATAAGAGCTGCTCCCCATCA 540

Qy    904 gacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaaaag 963
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAAG 600

Qy    964 tggataaatggtatgagctggggaatagttttttaataatatggcgggggagcagggtcc 1023
          ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db    601 TGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGGTCC 660

Qy   1024 actttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatgggaccc 1083
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAATGGGACCC 720

Qy   1084 gataaagtactggctgaacaggggccccggccctggagccaccgcataacttgccggtg 1143
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GATAAAGTACTGGCTGAACAGGGGCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTG 780

Qy   1144 ccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccactgga 1203
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACAGTACCACTGGA 840

Qy   1204 ttgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacagaga 1263
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGA 900

Qy   1264 ctcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgccact 1323
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCACT 960

Qy   1324 tcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaagaa 1383
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAA 1020

Qy   1384 agaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaataag 1443
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAG 1080

Qy   1444 cttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcccac 1503
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCAC 1140

Qy   1504 caacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtatttagta 1563
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTA 1200

Qy   1564 cctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacctca 1623
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCA 1260

Qy   1624 gtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtccccccgagtgtactac 1683
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GTCTTCAACCAATCCAAAGATTTGTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTAC 1320

Qy   1684 catcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaagagaa 1743
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGAA 1380

Qy   1744 cccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtaggaaca 1803
```

```
                                                
```

```
Query Match             59.6%;  Score 1763;  DB 14;  Length 8750;
Best Local Similarity   99.2%;  Pred. No. 0;
Matches 2913;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       1 tgctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgctct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5736 TGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTGCTCT 5795

Qy      61 acgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5796 ACGGGGGACCCCCCCCGTTGGTAGAAATTGCTTCTGTACATAGTGCTGATGTGCTGCTTT 5855

Qy     121 cccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 180
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5856 CCCAGCCTCTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAACGAGCGTGGAAGC 5915

Qy     181 agctccgggaggcctactcaggagagagagacttgcaagttccacatcgcttccaagttg 240
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||| |
Db    5916 AGCTCCGGGAGGCCTACTCAGGAGAAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTGG 5975

Qy     241 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggac 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    5976 GAGATTCAGTCTATGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGCC 6035

Qy     301 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6036 CTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGA 6095

Qy     361 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6096 TCCATGCATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGA 6155

Qy     421 ctgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactccc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6156 CTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCC 6215

Qy     481 caggccagtagtaaacgccttatagacagctcgaaccccccatagacctttatcccctacc 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    6216 CAGGCCAGTAGTAAACGCCTTATAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACC 6275

Qy     541 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6276 TGGCTGATTATTGACCCTGATACGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCT 6335

Qy     601 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6336 AGAGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAA 6395

Qy     661 agcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaa 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6396 AGCACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAA 6455

Qy     721 gagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacctcc 780
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    6456 GAGAAATACTGTGGGGGTTCTGAGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCC 6515

Qy     781 aacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtc 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6516 AACGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTC 6575

Qy     841 aattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctcccca 900
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    6576 AATTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCA 6635

Qy     901 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6636 TCAGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAA 6695

Qy     961 aagtggataaatggtatgagctggggaatagttttttataaatatggcgggggagcaggg 1020
           |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db    6696 AAGTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGG 6755

Qy    1021 tccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatggga 1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    6756 TCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGA 6815

Qy    1081 cccgataaagtactggctgaacaggggccccccggccctggagccaccgcataacttgccg 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6816 CCCGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCG 6875

Qy    1141 gtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccact 1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6876 GTGCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACT 6935

Qy    1201 ggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacag 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6936 GGATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAG 6995

Qy    1261 agactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgcc 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6996 AGACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCC 7055

Qy    1321 acttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaa 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7056 ACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAA 7115

Qy    1381 gaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaat 1440
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7116 GAAGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAAT 7175

Qy    1441 aagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcc 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7176 AAGCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCC 7235

Qy    1501 caccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtattta 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7236 CACCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTA 7295

Qy    1561 gtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacc 1620
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    7296 GTACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCTCC 7355

Qy    1621 tcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtcccccgagtgtac 1680
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    7356 TCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTAC 7415

Qy    1681 taccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaaga 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7416 TACCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGA 7475

Qy    1741 gaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtagga 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7476 GAACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGA 7535

Qy    1801 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7536 ACAGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTA 7595

Qy    1861 catgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagag 1920
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    7596 CATGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAG 7655

Qy    1921 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7656 TCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTT 7715

Qy    1981 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7716 CTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCAC 7775

Qy    2041 tcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagg 2100
           |||||||||||||||||||||||| |||||||||| ||| ||||||||||||||||||||
Db    7776 TCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGG 7835
```

```
Qy   2101 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 2160
          ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   7836 GAAAGAGAGGCTGACCAGGGGTGGTTTAAAGGATGGTTCAACAGGTCTCCTTGGATGACC 7895

Qy   2161 accctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttggg 2220
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   7896 ACCCTGCTTTCTGCTCTGACGGGACCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGG 7955

Qy   2221 ccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatc 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7956 CCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATC 8015

Qy   2281 atggtacttaggcaacagtaccaaggccttctgagccaaggagaaaactgacctctagcct 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8016 ATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCT 8075

Qy   2341 tcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaaaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8076 TCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTGGGGAATGAAAGGATGAAAA 8135

Qy   2401 tgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgcccccgaatt 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8136 TGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAGCTCTAAATGCCCCCGAATT 8195

Qy   2461 acagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggc 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8196 CCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGC 8255

Qy   2521 ctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggatt 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8256 CTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATT 8315

Qy   2581 ctgtaaaactgactggcaccatagaagaattgattacacattgacagccctagtgaccta 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8316 CTGTAAAACTGACTGGCACCATAGAAGAATTGATTACACATTGACAGCCCTAGTGACCTA 8375

Qy   2641 tctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagcta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8376 TCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGACCTCCGGAGCTA 8435

Qy   2701 ttttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattggtccatg 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   8436 TTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCACG 8495

Qy   2761 gagcgcgggctctcgatattttaaaatgattggtttgtgacgcacaggctttgttgtgaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8496 GAGCGCGGGCTCTCGATATTTTAAAATGATTGGTTTGTGACGCACAGGCTTTGTTGTGAA 8555

Qy   2821 ccccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcgtggtg 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8556 CCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTG 8615

Qy   2881 tacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatcaa 2936
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8616 TACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATCAA 8671
```

```
RESULT   5
AX002804
LOCUS       AX002804     3482 bp    DNA             PAT       21-AUG-2000
DEFINITION  Sequence 3 from Patent WO9853104.
ACCESSION   AX002804
VERSION     AX002804.1  GI:9885132
KEYWORDS    .
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 3482)
  AUTHORS   Stoye,J.P. and Weiss,R.A.
  TITLE     Detection of retroviral subtypes based upon envelope specific
            sequences
  JOURNAL   Patent: WO 9853104-A 3 26-NOV-1998;
            MEDICAL RES COUNCIL (GB); STOYE JONATHAN PAUL (GB)
FEATURES             Location/Qualifiers
     source          1. .3482
                     /organism="porcine endogenous retrovirus"
                     /db_xref="taxon:61673"
BASE COUNT       927 a     854 c     867 g     834 t
ORIGIN
```

Query Match 59.0%; Score 1743; DB 6; Length 3482;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2913; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

```
Qy      1 tgctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgctct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    548 TGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTGCTCT 607

Qy     61 acgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    608 ACGGGGGACCCCCCCCGTTGGTAGAAATTGCTTCTGTACATAGTGCTGATGTGCTGCTTT 667

Qy    121 cccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 180
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    668 CCCAGCCTCTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAACGAGCGTGGAAGC 727

Qy    181 agctccgggaggcctactcaggagagagagacttgcaagttccacatcgcttccaagttg 240
          |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||| |
Db    728 AGCTCCGGGAGGCCTACTCAGGAGAAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTGG 787

Qy    241 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggac 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    788 GAGATTCAGTCTATGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGCC 847

Qy    301 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    848 CTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGA 907

Qy    361 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    908 TCCATGCATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGA 967

Qy    421 ctgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactccc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    968 CTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCC 1027

Qy    481 caggccagtagtaaacgccttatagacagctcgaaccccccatagaccttttatccctacc 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   1028 CAGGCCAGTAGTAAACGCCTTATAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACC 1087

Qy    541 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1088 TGGCTGATTATTGACCCTGATACGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCT 1147

Qy    601 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1148 AGAGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAA 1207

Qy    661 agcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1208 AGCACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAA 1267

Qy    721 gagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacctcc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1268 GAGAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCC 1327

Qy    781 aacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtc 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1328 AACGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTC 1387

Qy    841 aattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctcccca 900
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1388 AATTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCA 1447
```

```
Qy    901 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1448 TCAGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAA 1507

Qy    961 aagtggataaatggtatgagctggggaatagtttttaataatatggcgggggagcaggg 1020
          ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db   1508 AAGTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGG 1567

Qy   1021 tccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatggga 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1568 TCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGA 1627

Qy   1081 cccgataaagtactggctgaacaggggcccccggccctggagccaccgcataacttgccg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1628 CCCGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCG 1687

Qy   1141 gtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccact 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   1688 GTGCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACT 1747

Qy   1201 ggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1748 GGATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAG 1807

Qy   1261 agactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgcc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1808 AGACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCC 1867

Qy   1321 acttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaa 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1868 ACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAA 1927

Qy   1381 gaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaat 1440
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1928 GAAGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAAT 1987

Qy   1441 aagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1988 AAGCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCC 2047

Qy   1501 caccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtattta 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2048 CACCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTA 2107

Qy   1561 gtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2108 GTACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACC 2167

Qy   1621 tcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtccccgagtgtac 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2168 TCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTAC 2227

Qy   1681 taccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaaga 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2228 TACCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGA 2287

Qy   1741 gaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtagga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2288 GAACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGA 2347

Qy   1801 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2348 ACAGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTA 2407

Qy   1861 catgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagag 1920
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   2408 CATGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAG 2467

Qy   1921 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2468 TCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTT 2527

Qy   1981 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2528 CTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCAC 2587

Qy   2041 tcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagg 2100
          |||||||||||||||||||||||||||||||| |||||||||| ||| ||||||||||||||
Db   2588 TCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGG 2647

Qy   2101 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2648 GAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACC 2707

Qy   2161 accctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttggg 2220
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   2708 ACCCTGCTTTCTGCTCTGACGGGACCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGG 2767

Qy   2221 ccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatc 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2768 CCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATC 2827

Qy   2281 atggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctagcct 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2828 ATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCT 2887

Qy   2341 tcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaaaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2888 TCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTGGGGAATGAAAGGATGAAAA 2947

Qy   2401 tgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgccccgaatt 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2948 TGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAGCTCTAAATGCCCCCGAATT 3007

Qy   2461 acagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggc 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3008 CCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGC 3067

Qy   2521 ctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggatt 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3068 CTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATT 3127

Qy   2581 ctgtaaaactgactggcaccatagaagaattgattacacattgacagccctagtgaccta 2640
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3128 CTGTAAAACCGACTGGCACCATAGAAGAATTGATTACACATTGACAGCCCTAGTGACCTA 3187

Qy   2641 tctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagcta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3188 TCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGACCTCCGGAGCTA 3247

Qy   2701 ttttaaaatgattggtccacggagcgcgggctctcgatatttttaaaatgattggtccatg 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   3248 TTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCACG 3307

Qy   2761 gagcgcgggctc-tcgatattttaaaatgattggtttgtgacgcacaggctttgttgtga 2819
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3308 GAGCGCGGGCTCTTCGATATTTTAAAATGATTGGTTTGTGACGCACAGGCTTTGTTGTGA 3367

Qy   2820 accccataaaagctgtcccgattccgcactcggggccgcagtcctctaccctgcgtggt 2879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3368 ACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGT 3427

Qy   2880 gtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 2934
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3428 GTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 3482
```

```
RESULT   6
PERENV2
LOCUS       PERENV2      3482 bp    RNA             VRL       14-OCT-1997
DEFINITION  Porcine endogenous retrovirus env gene, 3482 bp.
ACCESSION   Y12239
VERSION     Y12239.1  GI:2576324
KEYWORDS    env gene.
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
```

```
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 3482)
  AUTHORS   Le Tissier,P., Stoye,J.P., Takeuchi,Y., Patience,C. and Weiss,R.A.
  TITLE     Two sets of human-tropic pig retrovirus
  JOURNAL   Nature 389 (6652), 681-682 (1997)
  MEDLINE   97478526
REFERENCE   2  (bases 1 to 3482)
  AUTHORS   Stoye,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-1997) J.P. Stoye, National Institute for Medical
            Research, Virology, The Ridgeway Mill Hill, London, NW7 1AA, UK
  REMARK    revised by author 16-SEP-97
FEATURES             Location/Qualifiers
     source          1. .3482
                     /organism="porcine endogenous retrovirus"
                     /specific_host="Sus scrofa"
                     /db_xref="taxon:61673"
                     /note="PERV-B"
     gene            911. .2884
                     /gene="env"
     CDS             911. .2884
                     /gene="env"
                     /codon_start=1
                     /protein_id="CAA72928.1"
                     /db_xref="GI:2576325"
                     /db_xref="SPTREMBL:O41173"
                     /translation="MHPTLSWRHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSK
                     RLIDSSNPHRPLSLTWLIIDPDTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTP
                     PNLVRSYGFYCCPGTEKEKYCGGSGESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVN
                     SGPGKYKVMKLYKDKSCSPSDLDYLKISFTEKGKQENIQKWINGMSWGIVFYKYGGGA
                     GSTLTIRLRIETGTEPPVAVGPDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSN
                     GTTGLIPTNTPRNSPGVPVKTGQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYY
                     EGMAKEGKFNVTKEHRNQCTWGSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQ
                     ASENQYLVPGYNRWWACNTGLTPCVSTSVFNQSKDFCVMVQIVPRVVYYHPEEVVLDEY
                     DYRYNRPKREPVSLTLAVMLGLGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLR
                     ALEESVSNLEESLTSLSEVVLQNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDS
                     MSKLRERLERRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLIN
                     RFVAFVRERVSAVQIMVLRQQYQGLLSQGETDL"
BASE COUNT      927 a    854 c    867 g    834 t
ORIGIN
```

Query Match 59.0%; Score 1743; DB 14; Length 3482;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2913; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

```
Qy       1 tgctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgctct 60
Db     548 TGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTGCTCT 607

Qy      61 acgggggacccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 120
Db     608 ACGGGGGACCCCCCCCCGTTGGTAGAAATTGCTTCTGTACATAGTGCTGATGTGCTGCTTT 667

Qy     121 cccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 180
Db     668 CCCAGCCTCTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAACGAGCGTGGAAGC 727

Qy     181 agctccgggaggcctactcaggagagagagacttgcaagttccacatcgcttccaagttg 240
Db     728 AGCTCCGGGAGGCCTACTCAGGAGAAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTGG 787

Qy     241 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggac 300
Db     788 GAGATTCAGTCTATGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGCC 847

Qy     301 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 360
Db     848 CTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGA 907

Qy     361 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 420
Db     908 TCCATGCATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGA 967

Qy     421 ctgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactccc 480
Db     968 CTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCC 1027

Qy     481 caggccagtagtaaacgccttatagacagctcgaacccccatagacctttatcccctacc 540
Db    1028 CAGGCCAGTAGTAAACGCCTTATAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACC 1087

Qy     541 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 600
Db    1088 TGGCTGATTATTGACCCTGATACGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCT 1147

Qy     601 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 660
Db    1148 AGAGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAA 1207

Qy     661 agcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaa 720
Db    1208 AGCACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAA 1267

Qy     721 gagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacctcc 780
Db    1268 GAGAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCC 1327

Qy     781 aacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctccttttgtc 840
Db    1328 AACGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTC 1387

Qy     841 aattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctccccca 900
Db    1388 AATTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCA 1447

Qy     901 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 960
Db    1448 TCAGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAA 1507

Qy     961 aagtggataaatggtatgagctggggaatagttttttaataatatggcgggggagcaggg 1020
Db    1508 AAGTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGG 1567

Qy    1021 tccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatggga 1080
Db    1568 TCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGA 1627

Qy    1081 cccgataaagtactggctgaacaggggccccggccctggagccaccgcataacttgccg 1140
Db    1628 CCCGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCG 1687

Qy    1141 gtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccact 1200
Db    1688 GTGCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACT 1747

Qy    1201 ggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacag 1260
Db    1748 GGATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAG 1807

Qy    1261 agactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgcc 1320
Db    1808 AGACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCC 1867

Qy    1321 acttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaa 1380
Db    1868 ACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAA 1927

Qy    1381 gaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaat 1440
Db    1928 GAAGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAAT 1987

Qy    1441 aagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcc 1500
Db    1988 AAGCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCC 2047
```

```
Qy   1501 caccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtattta 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2048 CACCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTA 2107

Qy   1561 gtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2108 GTACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACC 2167

Qy   1621 tcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtcccccgagtgtac 1680
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   2168 TCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTAC 2227

Qy   1681 taccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaaga 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2228 TACCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGA 2287

Qy   1741 gaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtagga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2288 GAACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGA 2347

Qy   1801 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2348 ACAGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTA 2407

Qy   1861 catgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagag 1920
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   2408 CATGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAG 2467

Qy   1921 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2468 TCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTT 2527

Qy   1981 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2528 CTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCAC 2587

Qy   2041 tcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagg 2100
          |||||||||||||||||||||||||| |||||||||| ||| ||||||||||||||||||
Db   2588 TCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGG 2647

Qy   2101 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2648 GAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACC 2707

Qy   2161 accctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttggg 2220
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   2708 ACCCTGCTTTCTGCTCTGACGGGACCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGG 2767

Qy   2221 ccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatc 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2768 CCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATC 2827

Qy   2281 atggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctagcct 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2828 ATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCT 2887

Qy   2341 tcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaaaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2888 TCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTGGGGAATGAAAGGATGAAAA 2947

Qy   2401 tgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgccccgaattt 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2948 TGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAGCTCTAAATGCCCCCGAATT 3007

Qy   2461 acagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggc 2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3008 CCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGC 3067

Qy   2521 ctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggatt 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3068 CTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATT 3127

Qy   2581 ctgtaaaactgactggcaccatagaagaattgattacacattgacagccctagtgaccta 2640
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3128 CTGTAAAACCGACTGGCACCATAGAAGAATTGATTACACATTGACAGCCCTAGTGACCTA 3187

Qy   2641 tctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagcta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3188 TCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGACCTCCGGAGCTA 3247

Qy   2701 ttttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattggtccatg 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   3248 TTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCACG 3307

Qy   2761 gagcgcgggctc-tcgatattttaaaatgattggtttgtgacgcacaggctttgttgtga 2819
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3308 GAGCGCGGGCTCTTCGATATTTTAAAATGATTGGTTTGTGACGCACAGGCTTTGTTGTGA 3367

Qy   2820 accccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcgtggt 2879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3368 ACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGT 3427

Qy   2880 gtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 2934
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3428 GTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 3482
```

```
RESULT   7
PERY17013
LOCUS       PERY17013     7808 bp    DNA             VRL       14-APR-2000
DEFINITION  Porcine endogenous retrovirus gag, pol, env genes.
ACCESSION   Y17013
VERSION     Y17013.1  GI:3005985
KEYWORDS    env gene; gag gene; pol gene.
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 7808)
  AUTHORS   Czauderna,F., Fischer,N., Boller,K., Kurth,R. and Tonjes,R.R.
  TITLE     Establishment and characterization of molecular clones of porcine
            endogenous retroviruses replicating on human cells
  JOURNAL   J. Virol. 74 (9), 4028-4038 (2000)
  MEDLINE   20219394
REFERENCE   2  (bases 1 to 7808)
  AUTHORS   Toenjes,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1998) R.R. Toenjes, Paul-Ehrlich Institut,
            Paul-Ehrlich Str. 51-59, P.O.Box 1740, D-63207 Langen, FRG
FEATURES             Location/Qualifiers
     source          1. .7808
                     /organism="porcine endogenous retrovirus"
                     /db_xref="taxon:61673"
                     /tissue_type="kidney"
                     /cell_line="PK15"
     gene            260. .1834
                     /gene="gag"
     CDS             260. .1834
                     /gene="gag"
                     /codon_start=1
                     /protein_id="CAA76581.1"
                     /db_xref="GI:3005986"
                     /db_xref="SPTREMBL:O73504"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPSSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPRIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTAEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRLGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKAYYGRETEEEKEQRKEKEREEREERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKD"
     gene            2120. .4726
                     /gene="pol"
```

```
     CDS             2120. .4726
                     /gene="pol"
                     /codon_start=1
                     /protein_id="CAA76582.1"
                     /db_xref="GI:3005987"
                     /db_xref="SPTREMBL:O73505"
                     /translation="MGAQISFEQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQD
                     IQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQ
                     RLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLS
                     ALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNS
                     PTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYR
                     ASAKKAQICRREVTYLGYSLRGGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCR
                     LWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLY
                     VDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLT
                     LGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPE
                     ETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDG
                     THTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQR
                     GLLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQA
                     AQAVNLLPIIETPRAPEPRRQYTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEG
                     LEYVQQIHRLTHLGTKHLQQLVRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRIPP
                     GKRLRGSHPGAHWEVDFTEVKPAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKY
                     WRKFFRDLEYLR"
     gene            5294. .7267
                     /gene="env"
     CDS             5294. .7267
                     /gene="env"
                     /codon_start=1
                     /protein_id="CAA76583.1"
                     /db_xref="GI:3005988"
                     /db_xref="SPTREMBL:O73506"
                     /translation="MHPTLSWRHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSK
                     RLIDSSNPHRPLSLTWLIIDPDTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTP
                     PNLVRSYGFYCCPGTEKEKYCGGSEESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVN
                     SGPGKYKVMKLYKDKSCSPSDLDYLKISFTEKGKQENIQKWINGMSWGIVFYKYGGGA
                     GSTLTIRLRIETGTEPPVAVGPDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSN
                     GTTGLIPTNTPRNSPGVPVKTGQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYY
                     EGMAKEGKFNVTKEHRNQCTWGSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQ
                     ASENQYLVPGYNRWWACNTGLTPCVSSSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEY
                     DYRYNRPKREPVSLTLAVMLGLGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLR
                     ALEESVSNLEESLTSLSEVVLQNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDS
                     MSKLRERLERRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLIN
                     RFVAFVRERVGAVQIMVLRQQYQGLLSQGETDL"
BASE COUNT     2077 a   1971 c   2047 g   1713 t
ORIGIN
```

Query Match 56.0%; Score 1654; DB 14; Length 7808;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2854; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy      1 tgctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgctct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4931 TGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTGCTCT 4990

Qy     61 acgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4991 ACGGGGGACCCCCCCCCGTTGGTAGAAATTGCTTCTGTACATAGTGCTGATGTGCTGCTTT 5050

Qy    121 cccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 180
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5051 CCCAGCCTCTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAACGAGCGTGGAAGC 5110

Qy    181 agctccgggaggcctactcaggagagagagacttgcaagttccacatcgcttccaagttg 240
          |||||||||||||||||||||||  ||||||||||||||||||||||||||||||| |
Db   5111 AGCTCCGGGAGGCCTACTCAGGAGAAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTGG 5170

Qy    241 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggac 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   5171 GAGATTCAGTCTATGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGCC 5230

Qy    301 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5231 CTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGA 5290

Qy    361 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5291 TCCATGCATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGA 5350

Qy    421 ctgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactccc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5351 CTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCC 5410

Qy    481 caggccagtagtaaacgccttatagacagctcgaaccccccatagacctttatccctacc 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   5411 CAGGCCAGTAGTAAACGCCTTATAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACC 5470

Qy    541 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5471 TGGCTGATTATTGACCCTGATACGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCT 5530

Qy    601 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5531 AGAGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAA 5590

Qy    661 agcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5591 AGCACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAA 5650

Qy    721 gagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacctcc 780
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   5651 GAGAAATACTGTGGGGGTTCTGAGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCC 5710

Qy    781 aacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctccttttgtc 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5711 AACGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTC 5770

Qy    841 aattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctccccca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5771 AATTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCA 5830

Qy    901 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5831 TCAGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAA 5890

Qy    961 aagtggataaatggtatgagctggggaatagttttttaataatatggcgggggagcaggg 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5891 AAGTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGG 5950

Qy   1021 tccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatggga 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5951 TCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGA 6010

Qy   1081 cccgataaagtactggctgaacaggggccccggccctggagccaccgcataacttgccg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6011 CCCGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCG 6070

Qy   1141 gtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccact 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6071 GTGCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCAACGGTACCACT 6130

Qy   1201 ggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6131 GGATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAG 6190

Qy   1261 agactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgcc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6191 AGACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCC 6250

Qy   1321 acttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaa 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6251 ACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAA 6310

Qy   1381 gaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaat 1440
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6311 GAAGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAAT 6370
```

```
Qy   1441 aagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6371 AAGCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCC 6430

Qy   1501 caccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtattta 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6431 CACCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTA 6490

Qy   1561 gtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacc 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   6491 GTACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCTCC 6550

Qy   1621 tcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtcccccgagtgtac 1680
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   6551 TCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTAC 6610

Qy   1681 taccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaaga 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6611 TACCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGA 6670

Qy   1741 gaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtagga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6671 GAACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGA 6730

Qy   1801 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6731 ACAGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTA 6790

Qy   1861 catgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagag 1920
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   6791 CATGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAG 6850

Qy   1921 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6851 TCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTT 6910

Qy   1981 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6911 CTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCAC 6970

Qy   2041 tcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagg 2100
          |||||||||||||||||||||||| |||||||||| ||| ||||||||||||||||||||
Db   6971 TCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGG 7030

Qy   2101 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7031 GAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACC 7090

Qy   2161 accctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttggg 2220
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   7091 ACCCTGCTTTCTGCTCTGACGGGACCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGG 7150

Qy   2221 ccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatc 2280
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   7151 CCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAGTGGGTGCAGTCCAGATC 7210

Qy   2281 atggtacttaggcaacagtaccaaggccttctgagccaaggagaaaactgacctctagcct 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7211 ATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCT 7270

Qy   2341 tcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaaaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7271 TCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTGGGGAATGAAAGGATGAAAA 7330

Qy   2401 tgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgcccccgaatt 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7331 TGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAGCTCTAAATGCCCCCGAATT 7390

Qy   2461 acagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggc 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7391 CCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGC 7450

Qy   2521 ctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggatt 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7451 CTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGCTGACTAATCGCTTATCTGGATT 7510

Qy   2581 ctgtaaaactgactggcaccatagaagaattgattacacattgacagccctagtgaccta 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7511 CTGTAAAACTGACTGGCACCATAGAAGAATTGATTACACATTGACAGCCCTAGTGACCTA 7570

Qy   2641 tctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagcta 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7571 TCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGACCTCCGGAGCTA 7630

Qy   2701 ttttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattggtccatg 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   7631 TTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCACG 7690

Qy   2761 gagcgcgggctctcgatattttaaaatgattggtttgtgacgcacaggctttgttgtgaa 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7691 GAGCGCGGGCTCTCGATATTTTAAAATGATTGGTTTGTGACGCACAGGCTTTGTTGTGAA 7750

Qy   2821 ccccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcgtgg 2878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7751 CCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGG 7808
```

RESULT 8
PEN133816

```
LOCUS       PEN133816    8918 bp    DNA             VRL       07-DEC-2000
DEFINITION  Porcine endogenous retrovirus type C proviral gag, pol and env
            genes and LTR (class B, clone 33).
ACCESSION   AJ133816
VERSION     AJ133816.1  GI:6688945
KEYWORDS    env gene; env protein; gag gene; gag protein; pol gene; pol
            protein.
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 8918)
  AUTHORS   Czauderna,F., Fischer,N., Boller,K., Kurth,R. and Tonjes,R.R.
  TITLE     Establishment and characterization of molecular clones of porcine
            endogenous retroviruses replicating on human cells
  JOURNAL   J. Virol. 74 (9), 4028-4038 (2000)
  MEDLINE   20219394
REFERENCE   2  (bases 1 to 8918)
  AUTHORS   Toenjes,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-1999) Toenjes R.R., Medical Biotechnology,
            Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, Hessen, D-63225
            Langen, GERMANY
FEATURES             Location/Qualifiers
     source          1. .8918
                     /organism="porcine endogenous retrovirus"
                     /proviral
                     /strain="type C"
                     /db_xref="taxon:61673"
                     /clone="33"
                     /note="grown in PK15 (ATCC No. CCL-33) cell line"
     LTR             1. .707
     misc_feature    545
                     /note="cap site"
     primer_bind     710. .727
                     /note="gly4"
     gene            1154. .2725
                     /gene="gag"
     mat_peptide     1154. .2725
                     /gene="gag"
                     /function="retroviral capsid"
                     /product="Gag protein"
     CDS             join(1154. .2725,2729. .6313,6189. .8162)
                     /note="viral polypeptide"
                     /codon_start=1
                     /protein_id="CAB65339.1"
                     /db_xref="GI:6688946"
```

```
                     /db_xref="SPTREMBL:Q9Q1X5"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPSSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPRIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREEREERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLTLEEDKDGRRGSDPLPEPRVTLKVEGQPVEFLVDTGAEHSVLLQPLGKL
                     KEKKSWVMGATGQRQYPWTTRRTVDLAVGRVTHSFLVIPECPVPLLGRDLLTKMGAQI
                     SFEQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMG
                     LAKQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPL
                     LPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFF
                     CLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQH
                     PQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYS
                     LRGGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKE
                     KGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPW
                     RRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQ
                     PPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETG
                     VRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAE
                     LMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSL
                     LEALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPIIETPKAPEPR
                     RQYTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEGLEYVQQIHRLTHLGTKHLQ
                     QLVRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRIPPGKRLRGSHPGAHWEVDFTE
                     VKPAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKILEEIFPRFGIPKVIGSDNG
                     PAFVAQVSQGLAKILGIDWKLHCAYRPQSSGQVERMNRTIKETLTKLTTETGINDWIA
                     LLPFVLFRVRNTPGQFGLTPYELLYGGPPPLVEIASVHSADVLLSQPLFSRLKALEWV
                     RQRAWKQLREAYSGEGDLQVPHRFQVGDSVYVRRHRAGNLETRWKGPYLVLLTTPTAV
                     KVEGISTWIHTSHVKLAPPPDSGWRAEKTENPLKLRLHRLVPYSNNNSPGQIHPTLSW
                     RHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSKRLIDSSNPHRPLSLTWLIIDP
                     DTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTPPNLVRSYGFYCCPGTEKEKYC
                     GDSGESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVNSGPGKYKVMKLYKDKSCSPSD
                     LDYLKISFTEKGKQENIQKWINGMSWGIVFYKYGGGAGSTLTIRLRIETGTEPPVAVG
                     PDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQSPSNGTTGLIPTNTPRNSPGVPVKT
                     GQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYYEGMAKEGKFNVTKEHRNQCTW
                     GSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQASENQYLVPGYNRWWACNTGL
                     TPCVSTSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEYDYRYNRPKREPVSLTLAVMLG
                     LGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLRALEESVSNLEESLTSLSEVVL
                     QNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDSMSKLRERLERRRREREADQGW
                     FEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLINRFVAFVRERVSAVQIMVLRQQ
                     YQGLLSQGETDL"
     mat_peptide     2729. .6313
                     /gene="pol"
                     /function="retroviral protease and polymerase"
                     /product="Pol protein"
     gene            2729. .6313
                     /gene="pol"
     mat_peptide     6189. .8162
                     /gene="env"
                     /function="retroviral envelope"
                     /product="Env protein"
     gene            6189. .8162
                     /gene="env"
     LTR             8212. .8918
     polyA_signal    8811. .8816
BASE COUNT     2327 a    2245 c    2334 g    2012 t
ORIGIN
```

Query Match 55.3%; Score 1636; DB 14; Length 8918;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2736; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy       1 tgctttttagggttaggaacacccctggacagtttgggctgacccccatgaattgctct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5826 TGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTGCTCT 5885

Qy      61 acgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5886 ACGGGGGACCCCCCCCGTTGGTAGAAATTGCTTCTGTACATAGTGCTGATGTGCTGCTTT 5945

Qy     121 cccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 180
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5946 CCCAGCCTCTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAACGAGCGTGGAAGC 6005

Qy     181 agctccgggaggcctactcaggagagagagacttgcaagttccacatcgcttccaagttg 240
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||| |
Db    6006 AGCTCCGGGAGGCCTACTCAGGAGAAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTGG 6065

Qy     241 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggac 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6066 GAGATTCAGTCTATGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGCC 6125

Qy     301 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6126 CTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGA 6185

Qy     361 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 420
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6186 TCCATACATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGA 6245

Qy     421 ctgagaatcccctaagcttcgcctccatcgcctggttccttactctaacaataactccc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6246 CTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCC 6305

Qy     481 caggccagtagtaaacgccttatagacagctcgaacccccatagacctttatcccctacc 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    6306 CAGGCCAGTAGTAAACGCCTTATAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACC 6365

Qy     541 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6366 TGGCTGATTATTGACCCTGATACGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCT 6425

Qy     601 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6426 AGAGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAA 6485

Qy     661 agcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaa 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6486 AGCACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAA 6545

Qy     721 gagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacctcc 780
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    6546 GAGAAATACTGTGGGGATTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCC 6605

Qy     781 aacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtc 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6606 AACGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTC 6665

Qy     841 aattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctcccca 900
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    6666 AATTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCA 6725

Qy     901 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6726 TCAGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAA 6785

Qy     961 aagtggataaatggtatgagctggggaatagttttttaataatatggcgggggagcaggg 1020
           |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db    6786 AAGTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGG 6845

Qy    1021 tccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatggga 1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    6846 TCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGA 6905

Qy    1081 cccgataaagtactggctgaacaggggccccccggccctggagccaccgcataacttgccg 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6906 CCCGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCG 6965

Qy    1141 gtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccact 1200
           |||||||||||||||||||||||||||||||||||| |||||||||| ||||||||
Db    6966 GTGCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGTCGCCTAGCAACGGTACCACT 7025

Qy    1201 ggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacag 1260
```

```
Db   7026 GGATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAG 7085

Qy   1261 agactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgcc 1320
Db   7086 AGACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCC 7145

Qy   1321 acttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaa 1380
Db   7146 ACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAA 7205

Qy   1381 gaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaat 1440
Db   7206 GAAGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAAT 7265

Qy   1441 aagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcc 1500
Db   7266 AAGCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCC 7325

Qy   1501 caccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtattta 1560
Db   7326 CACCAACACCTTTGCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTA 7385

Qy   1561 gtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacc 1620
Db   7386 GTACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACC 7445

Qy   1621 tcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtcccccgagtgtac 1680
Db   7446 TCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTAC 7505

Qy   1681 taccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaaga 1740
Db   7506 TACCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGA 7565

Qy   1741 gaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtagga 1800
Db   7566 GAACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGA 7625

Qy   1801 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 1860
Db   7626 ACAGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTA 7685

Qy   1861 catgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagag 1920
Db   7686 CATGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAG 7745

Qy   1921 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 1980
Db   7746 TCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTT 7805

Qy   1981 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 2040
Db   7806 CTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCAC 7865

Qy   2041 tcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagg 2100
Db   7866 TCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGG 7925

Qy   2101 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 2160
Db   7926 GAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACC 7985

Qy   2161 accctgctttctgctctgacgggggccctagtagtcctgctcctgttacttacagttggg 2220
Db   7986 ACCCTGCTTTCTGCTCTGACGGGACCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGG 8045

Qy   2221 ccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatc 2280
Db   8046 CCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATC 8105

Qy   2281 atggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctagcct 2340
Db   8106 ATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCT 8165

Qy   2341 tcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaaaa 2400
Db   8166 TCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTGGGGAATGAAAGGATGAAAA 8225

Qy   2401 tgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgcccccgaatt 2460
Db   8226 TGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAGCTCTAAATGCCCCCGAATT 8285

Qy   2461 acagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggc 2520
Db   8286 CCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGC 8345

Qy   2521 ctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggatt 2580
Db   8346 CTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATT 8405

Qy   2581 ctgtaaaactgactggcaccatagaagaattgattacacattgacagccctagtgaccta 2640
Db   8406 CTGTAAAACTGACTGGCACCATAGAAGAATTGATTACACATTGACAGCCCTAGTGACCTA 8465

Qy   2641 tctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagcta 2700
Db   8466 TCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGACCTCCGGAGCTA 8525

Qy   2701 ttttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattggtcca 2758
Db   8526 TTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCA 8583
```

```
RESULT   9
PEN293657
LOCUS       PEN293657     8763 bp    DNA             VRL       30-MAY-2001
DEFINITION  Porcine endogenous type C retrovirus proviral gag gene, pol gene
            and env gene, class B, clone 213.
ACCESSION   AJ293657
VERSION     AJ293657.1  GI:14275841
KEYWORDS    env gene; envelope; gag gene; group specific antigen; pol gene;
            polymerase; polyprotein.
SOURCE      porcine endogenous type C retrovirus.
  ORGANISM  porcine endogenous type C retrovirus
            Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
REFERENCE   1  (bases 1 to 8763)
  AUTHORS   Krach,U., Fischer,N., Czauderna,F. and Tonjes,R.R.
  TITLE     Comparison of replication-competent molecular clones of porcine
            endogenous retrovirus class a and class b derived from pig and
            human cells
  JOURNAL   J. Virol. 75 (12), 5465-5472 (2001)
  MEDLINE   21256017
REFERENCE   2  (bases 1 to 8763)
  AUTHORS   Toenjes,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2000) Toenjes R.R., Medical Biotechnology,
            Paul-Ehrlich-Institut, Paul-Ehrlich-Str. 51-95, Langen, 63225,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .8763
                     /organism="porcine endogenous type C retrovirus"
                     /proviral
                     /specific_host="Sus scrofa"
                     /db_xref="taxon:70540"
                     /clone="213"
                     /cell_line="PK15"
                     /note="class B"
     LTR             1. .630
     misc_feature    467
                     /note="Cap site"
     primer_bind     633. .650
     gene            1077. .2651
                     /gene="gag"
     CDS             1077. .2651
                     /gene="gag"
```

```
                     /codon_start=1
                     /product="group specific antigen"
                     /protein_id="CAC39625.1"
                     /db_xref="GI:14275842"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPGSHPDQEPYILTWQDLAEEPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPGIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREEREERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKD"
     gene            2652. .6239
                     /gene="pol"
     CDS             <2652. .6239
                     /gene="pol"
                     /note="complete pol ORF, gag-pol is expressed as a
                     polyprotein"
                     /citation=[1]
                     /codon_start=1
                     /product="polymerase"
                     /protein_id="CAC39626.1"
                     /db_xref="GI:14275843"
                     /translation="GRRGSDPLPEPRVTLKVEGQPVEFLVDTGAEHSVLLQPLGKLKE
                     KKSWVMGATGQRQYPWTTRRTVDLAVGRVTHSFLVIPECPVPLLGRDLLTKMGAQISF
                     EQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLA
                     KQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLP
                     VRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCL
                     RLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTFFDEALHRDLANFRIQHPQ
                     VTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLR
                     GGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKG
                     EFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRR
                     PVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPP
                     DRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVR
                     KDLTDIPLTGEVLTWFTDGSSYVVEGKKMAGAAVVDGTRTIWASSLPEGTSAQKAELM
                     ALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLE
                     ALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPIIETPKAPEPRRQ
                     YTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEGLEYVQQIHRLTHLGTKHLQQL
                     VRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRIPPGKRLRGSHPGAHWEVDFTEVK
                     PAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKILEEIFPRFGIPKVIGSDNGPA
                     FVAQVSQGLAKILGIDWKLHCAYRPQSSGQVERMNRTIKETLTKLTTETGINDWIALL
                     PFVLFRVRNTPGQPGLTPYELLYGGPPPLVEIASVHSADVLLSQPLFSRLKALEWVRQ
                     RAWKQLREAYSGEGDLQVPHRFQVGDSVYVRRHHAGNLETRWKGPYLVLLTTPTAVKV
                     EGISTWIHASHVKLAPPPDSGWRAEKTENPLKLRLHRLVPYSNNNSPGQ"
     gene            6112. .8085
                     /gene="env"
     CDS             6112. .8085
                     /gene="env"
                     /codon_start=1
                     /product="envelope"
                     /protein_id="CAC39627.1"
                     /db_xref="GI:14275844"
                     /translation="MHPTLSWRHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSK
                     RLIDSSNPHRPLSLTWLIIDPDTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTP
                     PNLVRSYGFYCCPGTEKEKYCGGSGESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVN
                     SGPGKYKVMKLYKDKSCSPSDLDYLKISFTEKGKQENIQKWINGMSWGIVFYKYGGGA
                     GSTLTIRLRIETGTEPPVAVGPDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSX
                     STTGLIPTNTPRNSPGVPVKTGQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYY
                     EGMAKEGKFNVTKEHRNQCTWGSRNKLTLTEVSGKGTCIGKAPPSHQHLCNSTVVYEQ
                     ASENQYLVPGYNRWWACNTGLTPCVSTSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEY
                     DYRYNRPKREPVSLTLAVMLGLGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLR
                     ALEESVSNLEESLTSLSEVVLQNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDS
                     MSKLRERLERRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLIN
                     RFVAFVRERVSAVQIMVLRQQYQGLLSQGETDL"
     LTR             8133. .8763
     polyA_signal    8655. .8660
BASE COUNT     2301 a   2200 c   2287 g   1974 t      1 others
ORIGIN
```

Query Match 53.2%; Score 1572; DB 14; Length 8763;

Best Local Similarity 99.2%; Pred. No. 0;
Matches 2622; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy      1 tgctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgctct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5749 TGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTGCTCT 5808

Qy     61 acgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5809 ACGGGGGACCCCCCCCGTTGGTAGAAATTGCTTCTGTACATAGTGCTGATGTGCTGCTTT 5868

Qy    121 cccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5869 CCCAGCCTTTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAACGAGCGTGGAAGC 5928

Qy    181 agctccgggaggcctactcaggagagagacttgcaagttccacatcgcttccaagttg 240
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||| |
Db   5929 AGCTCCGGGAGGCCTACTCAGGAGAAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTGG 5988

Qy    241 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggac 300
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||| |
Db   5989 GAGATTCAGTCTATGTTAGACGCCACCATGCAGGAAACCTCGAGACTCGGTGGAAGGGCC 6048

Qy    301 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6049 CTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGA 6108

Qy    361 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6109 TCCATGCATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGA 6168

Qy    421 ctgagaatcccctaagcttcgcctccatcgcctggttccttactctaacaataactccc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6169 CTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCC 6228

Qy    481 caggccagtagtaaacgccttatagacagctcgaacccccatagacctttatcccctacc 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   6229 CAGGCCAGTAGTAAACGCCTTATAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACC 6288

Qy    541 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6289 TGGCTGATTATTGACCCTGATACGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCT 6348

Qy    601 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6349 AGAGGCACCTGGTGGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAA 6408

Qy    661 agcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6409 AGCACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAA 6468

Qy    721 gagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacctcc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6469 GAGAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCC 6528

Qy    781 aacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtc 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6529 AACGATGGAGACTGGAAATGGCCGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTC 6588

Qy    841 aattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctcccca 900
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||| ||||
Db   6589 AATTCCGGCCCGGGCAAGTACAAAGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCA 6648

Qy    901 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6649 TCAGACTTAGATTATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAA 6708

Qy    961 aagtggataaatggtatgagctggggaatagttttttaataatatggcgggggagcaggg 1020
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   6709 AAGTGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGGAGCAGGG 6768

Qy   1021 tccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatggga 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
```

```
Db   6769 TCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAACCCCCTGTGGCAGTGGGA 6828

Qy   1081 cccgataaagtactggctgaacaggggcccccggccctggagccaccgcataacttgccg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6829 CCCGATAAAGTACTGGCTGAACAGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCG 6888

Qy   1141 gtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccact 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   6889 GTGCCCCAATTAACCTCGCTGCGGCCTGACATAACACAGCCGCCTAGCANCAGTACCACT 6948

Qy   1201 ggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6949 GGATTGATTCCTACCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAG 7008

Qy   1261 agactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgcc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7009 AGACTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCC 7068

Qy   1321 acttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaa 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7069 ACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAA 7128

Qy   1381 gaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaat 1440
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7129 GAAGGAAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAAT 7188

Qy   1441 aagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctccccccatcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7189 AAGCTTACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCC 7248

Qy   1501 caccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtattta 1560
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   7249 CACCAACACCTTTGCAATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTA 7308

Qy   1561 gtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7309 GTACCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACC 7368

Qy   1621 tcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtcccccgagtgtac 1680
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   7369 TCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTAC 7428

Qy   1681 taccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaaga 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7429 TACCATCCTGAGGAAGTGGTCCTTGATGAATATGACTATCGGTATAACCGACCAAAAAGA 7488

Qy   1741 gaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtagga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7489 GAACCCGTATCCCTTACCCTAGCTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGA 7548

Qy   1801 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7549 ACAGGGACAGCTGCCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTA 7608

Qy   1861 catgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagag 1920
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   7609 CATGCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCTAGAAGAG 7668

Qy   1921 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7669 TCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTT 7728

Qy   1981 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7729 CTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCAC 7788

Qy   2041 tcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagg 2100
          |||||||||||||||||||||||| |||||||||| ||| ||||||||||||||||||||
Db   7789 TCAGGAGCCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGG 7848

Qy   2101 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7849 GAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACC 7908

Qy   2161 accctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttggg 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7909 ACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGG 7968

Qy   2221 ccttgcttaattaataggtttgttgccttttgttagagaacgagtgagtgcagtccagatc 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7969 CCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATC 8028

Qy   2281 atggtacttaggcaacagtaccaaggccttctgagccaaggagaaaactgacctctagcct 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8029 ATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCT 8088

Qy   2341 tcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaaaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8089 TCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTGGGGAATGAAAGGATGAAAA 8148

Qy   2401 tgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgccccgaatt 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8149 TGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAGCTCTAAATGCCCCGAATT 8208

Qy   2461 acagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggc 2520
          |||||||||||||||||||||||||||||||||||||| |||||||||| |||
Db   8209 CCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTGCTATTGTTCCAAGGC 8268

Qy   2521 ctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggatt 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8269 CTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATT 8328

Qy   2581 ctgtaaaactgactggcaccatagaagaattgattacacattgacagccctagtgaccta 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8329 CTGTAAAACTGACTGGCACCATAGAAGAATTGATTACACATTGACAGCCCTAGTGACCTA 8388

Qy   2641 tct 2643
          |||
Db   8389 TCT 8391
```

```
RESULT  10
AF014162
LOCUS       AF014162     1971 bp    mRNA            VRL       01-JUN-1998
DEFINITION  Porcine endogenous retrovirus envelope protein (env) mRNA, complete
            cds.
ACCESSION   AF014162
VERSION     AF014162.1  GI:3169735
KEYWORDS    .
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 1971)
  AUTHORS   Haworth,C., Galbraith,D.N., Lees,G.M. and Smith,K.T.
  TITLE     Porcine endogenous retrovirus (PoEV) Env sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1971)
  AUTHORS   Haworth,C., Galbraith,D.N., Lees,G.M. and Smith,K.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1997) Q-One Biotech Ltd., Todd Campus, Acre Rd.,
            Glasgow G20 0XA, UK
FEATURES             Location/Qualifiers
     source          1. .1971
                     /organism="porcine endogenous retrovirus"
                     /db_xref="taxon:61673"
                     /cell_line="derived from cell line PK15"
                     /note="PoEV"
     gene            1. .1971
                     /gene="env"
     CDS             1. .1971
                     /gene="env"
                     /codon_start=1
                     /product="envelope protein"
                     /protein_id="AAC17948.1"
                     /db_xref="GI:3169736"
```

```
                     /translation="MHPTLSRRHLPTRGGEPKRLRIPLSFASIAWFLTLTITPQASSK
                     RLIDSSNPHRPLSLTWLIIDPDTGVTVNSTRGVAPRGTWWPELHFCLRLINPAVKSTP
                     PNLVRSYGFYCCPGTEKEKYCGGSGESFCRRWSCVTSNDGDWKWPISLQDRVKFSFVN
                     SGPGKYKMMKLYKDKSCSPSDLDYLKISFTERKTGKYSKVDKWYELGNSFLLYGGGAG
                     STLTIRLRIETGTEPPVAMGPDKVLAEQGPPALEPPHNLPVPQLTSLRPDITQPPSNS
                     TTGLIPTNTPRNSPGVPVKTGQRLFSLIQGAFQAINSTDPDATSSCWLCLSSGPPYYE
                     GMAKERKFNVTKEHRNQCTWGSRNKLTLTEVSGKGTCIGKAPPSHQHLCYSTVVYEQA
                     SENQYLVPGYNRWWACNTGLTPCVSTSVFNQSKDFCVMVQIVPRVYYHPEEVVLDEYD
                     YRYNRPKREPVSLTLAVMLGLGTAVGVGTGTAALITGPQQLEKGLGELHAAMTEDLRA
                     LKESVSNLEESLTSLSEVVLQNRRGLDLLFLREGGLCAALKEECCFYVDHSGAIRDSM
                     NKLRKKLERRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLINR
                     FVAFVRERVSAVQIMVLRQQYQGLLSQGETDL"
BASE COUNT      530 a    490 c    482 g    469 t
ORIGIN
```

Query Match 50.7%; Score 1499; DB 14; Length 1971;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

```
Qy    383 ggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcg 442
Db     20 GGCGCCACCTCCCGACTCGGGGTGGAGAGCCGAAAAGACTGAGAATCCCCTTAAGCTTCG 79

Qy    443 cctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttа 502
Db     80 CCTCCATCGCCTGGTTCCTTACTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTA 139

Qy    503 tagacagctcgaaccccccatagacctttatcccctacctggctgattattgaccctgata 562
Db    140 TAGACAGCTCGAACCCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGATA 199

Qy    563 cgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac 622
Db    200 CGGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGTGGCCTGAAC 259

Qy    623 tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtcc 682
Db    260 TGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGCACACCTCCCAACCTAGTCC 319

Qy    683 gtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctg 742
Db    320 GTAGTTATGGGTTCTATTGCTGCCCAGGCACAGAGAAAGAGAAATACTGTGGGGGTTCTG 379

Qy    743 gggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggc 802
Db    380 GGGAATCCTTCTGTAGGAGATGGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGC 439

Qy    803 cgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtaca 862
Db    440 CGATCTCTCTCCAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACA 499

Qy    863 aaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaaga 922
Db    500 AAATGATGAAACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATTATCTAAAGA 559

Qy    923 taagtttcactgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagct 982
Db    560 TAAGTTTCACTG--AAAGGAAAACAGGAAAATATTCAAAAGTGGATAAATGGTATGAGCT 617

Qy    983 ggggaatagtttttttaataatatggcgggggagcagggtccactttaaccattcgccttа 1042
Db    618 GGGGAATAGTTTTTT-ATTATATGGCGGGGGAGCAGGGTCCACTTTAACCATTCGCCTTA 676

Qy   1043 ggatagagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaac 1102
Db    677 GGATAGAGACGGGGACAGAACCCCCTGTGGCAATGGGACCCGATAAAGTACTGGCTGAAC 736

Qy   1103 aggggccccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgc 1162
Db    737 AGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGC 796

Qy   1163 ggcctgacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgc 1222
Db    797 GGCCTGACATAACACAGCCGCCTAGCAACAGTACCACTGGATTGATTCCTACCAACACGC 856

Qy   1223 ctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagg 1282
Db    857 CTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGG 916

Qy   1283 gagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtc 1342
Db    917 GAGCTTTCCAAGCCATCAACTCCACCGACCCTGATGCCACTTCTTCTTGTTGGCTTTGTC 976

Qy   1343 tatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgacca 1402
Db    977 TATCCTCAGGGCCTCCTTATTATGAGGGGATGGCTAAAGAAAGAAAATTCAATGTGACCA 1036

Qy   1403 aagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagttt 1462
Db   1037 AAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTT 1096

Qy   1463 ccgggaaggggacatgcataggaaaagctcccccatcccaccaacaccttтgctatagta 1522
Db   1097 CCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCTATAGTA 1156

Qy   1523 ctgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggt 1582
Db   1157 CTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTACCTGGTTATAACAGGTGGT 1216

Qy   1583 gggcatgcaatactgggttaacccccтgtgtttccacctcagtcttcaaccaatccaaag 1642
Db   1217 GGGCATGCAATACTGGGTTAACCCCCTGTGTTTCCACCTCAGTCTTCAACCAATCCAAAG 1276

Qy   1643 atttgtgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtcc 1702
Db   1277 ATTTCTGTGTCATGGTCCAAATCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCC 1336

Qy   1703 ttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctag 1762
Db   1337 TTGATGAATATGACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAG 1396

Qy   1763 ctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatca 1822
Db   1397 CTGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTGCCCTGATCA 1456

Qy   1823 caggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatc 1882
Db   1457 CAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACATGCGGCCATGACAGAAGATC 1516

Qy   1883 tccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaag 1942
Db   1517 TCCGAGCCTTAAAGGAGTCTGTTAGCAACCTAGAAGAGTCCCTGACTTCTTTGTCTGAAG 1576

Qy   1943 tggttctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtg 2002
Db   1577 TGGTTCTACAGAACCGGAGGGGATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTG 1636

Qy   2003 cagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactcca 2062
Db   1637 CAGCCTTAAAAGAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCA 1696

Qy   2063 tgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggt 2122
Db   1697 TGAACAAGCTTAGAAAAAAGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGT 1756

Qy   2123 ggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacgg 2182
Db   1757 GGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGG 1816

Qy   2183 ggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttg 2242
Db   1817 GGCCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTG 1876

Qy   2243 ttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtacc 2302
```

```
Db    1877 TTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACC 1936

Qy    2303 aaggccttctgagccaaggagaaactgacctctag 2337
           |||||||||||||||||||||||||||||||||||
Db    1937 AAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAG 1971


RESULT  11
AF038601
LOCUS       AF038601     7333 bp    mRNA            MAM       21-MAY-1998
DEFINITION  Sus scrofa porcine endogenous retrovirus ERV-PK15 mRNA, complete
            sequence.
ACCESSION   AF038601
VERSION     AF038601.1  GI:3133304
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 7333)
  AUTHORS   Akiyoshi,D.E., Denaro,M., Zhu,H., Greenstein,J.L., Banerjee,P. and
            Fishman,J.A.
  TITLE     Identification of a full-length cDNA for an endogenous retrovirus
            of miniature swine
  JOURNAL   J. Virol. 72 (5), 4503-4507 (1998)
  MEDLINE   98216827
REFERENCE   2  (bases 1 to 7333)
  AUTHORS   Fishman,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1997) Infectious Disease Unit, Massachusetts
            General Hospital, 55 Fruit Street, Boston, MA 02114, USA
REFERENCE   3  (bases 1 to 7333)
  AUTHORS   Fishman,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-1998) Infectious Disease Unit, Massachusetts
            General Hospital, 55 Fruit Street, Boston, MA 02114, USA
  REMARK    Sequence update by submitter
COMMENT     On May 16, 1998 this sequence version replaced gi:3116445.
FEATURES             Location/Qualifiers
     source          1. .7333
                     /organism="Sus scrofa"
                     /strain="miniature swine"
                     /db_xref="taxon:9823"
                     /cell_line="PK-15"
                     /note="porcine endogenous retrovirus ERV-PK15; infective
                     virus but represents a naturally occurring transcript"
     LTR             1. .578
     CDS             598. .2172
                     /codon_start=1
                     /product="gag protein"
                     /protein_id="AAC16766.1"
                     /db_xref="GI:3116446"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPGSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPRIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     KKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREEREERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKD"
     CDS             2320. .3552
                     /note="truncated protein"
                     /codon_start=1
                     /product="pol protein"
                     /protein_id="AAC16767.1"
                     /db_xref="GI:3133305"
                     /translation="MGATGQRQYPWTTRRTVDLGVGRVTHSFLVIPECPVPLLGRDLL
                     TKMGAQISFEQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAW
                     AETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQ
                     SPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVL
                     DLKDAFFCLRLHPTSQPLFTFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDL
                     ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRRE
                     VTYLGYSLRGGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAP
                     LYPLTKEKGGCLPQQGGK"
     CDS             4803. .6725
                     /note="env protein; N terminus deleted"
                     /codon_start=1
                     /pseudo
     LTR             6726. .7333
BASE COUNT     1984 a   1788 c   1885 g   1676 t
ORIGIN


  Query Match           27.5%;  Score 813;  DB 4;  Length 7333;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches  863;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    2082 gttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaa 2141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6470 GTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAA 6529

Qy    2142 caggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgct 2201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6530 CAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTGCT 6589

Qy    2202 cctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacg 2261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6590 CCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACG 6649

Qy    2262 agtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagg 2321
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6650 AGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGG 6709

Qy    2322 agaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtg 2381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6710 AGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTG 6769

Qy    2382 gggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaag 2441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6770 GGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAG 6829

Qy    2442 ctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtagaaggtcaca 2501
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    6830 CTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACA 6889

Qy    2502 cttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttga 2561
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6890 CTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGA 6949

Qy    2562 ctaatcgcttatctggattctgtaaaactgactggcaccatagaagaattgattacacat 2621
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6950 CTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACCATAGAAGAATTGATTACACAT 7009

Qy    2622 tgacagccctagtgacctatctcaactgcaatctgtcactctgcccaggagcccacgcag 2681
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7010 TGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAG 7069

Qy    2682 atgcggacctccggagctattttaaaatgattggtccacggagcgcgggctctcgatatt 2741
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7070 ATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATT 7129

Qy    2742 ttaaaatgattggtccatggagcgcgggctctcgatattttaaaatgattggtttgtgac 2801
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7130 TTAAAATGATTGGTCCATGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTTTGTGAC 7189

Qy    2802 gcacaggctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgcagt 2861
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7190 GCACAGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGT 7249

Qy    2862 cctctacccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctct 2921
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7250 CCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCT 7309

Qy    2922 tgctgtttgcatcaaaaaaaaaaa 2945
```

```
          ||||||||||||||||||||||||
Db   7310 TGCTGTTTGCATCAAAAAAAAAAA 7333


RESULT  12
AR130474
LOCUS       AR130474     7333 bp    DNA             PAT       16-MAY-2001
DEFINITION  Sequence 2 from patent US 6190861.
ACCESSION   AR130474
VERSION     AR130474.1  GI:14118799
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7333)
  AUTHORS   Fishman,J.A.
  TITLE     Molecular sequences of swine retroviruses method of using
  JOURNAL   Patent: US 6190861-A 2 20-FEB-2001;
FEATURES             Location/Qualifiers
     source          1. .7333
                     /organism="unknown"
BASE COUNT     1984 a   1788 c   1885 g   1676 t
ORIGIN


  Query Match             27.5%;  Score 813;  DB 6;  Length 7333;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches  863;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   2082 gttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaa 2141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6470 GTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAA 6529

Qy   2142 caggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgct 2201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6530 CAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTGCT 6589

Qy   2202 cctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacg 2261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6590 CCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACG 6649

Qy   2262 agtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagg 2321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6650 AGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGG 6709

Qy   2322 agaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtg 2381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6710 AGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTG 6769

Qy   2382 gggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaag 2441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6770 GGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAG 6829

Qy   2442 ctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtagaaggtcaca 2501
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   6830 CTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACA 6889

Qy   2502 cttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttga 2561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6890 CTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGA 6949

Qy   2562 ctaatcgcttatctggattctgtaaaactgactggcaccatagaagaattgattacacat 2621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6950 CTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACCATAGAAGAATTGATTACACAT 7009

Qy   2622 tgacagccctagtgacctatctcaactgcaatctgtcactctgcccaggagcccacgcag 2681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7010 TGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAG 7069

Qy   2682 atgcggacctccggagctattttaaaatgattggtccacggagcgcgggctctcgatatt 2741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7070 ATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATT 7129

Qy   2742 ttaaaatgattggtccatggagcgcgggctctcgatattttaaaatgattggtttgtgac 2801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7130 TTAAAATGATTGGTCCATGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTTTGTGAC 7189

Qy   2802 gcacaggctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgcagt 2861
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7190 GCACAGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGT 7249

Qy   2862 cctctacccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctct 2921
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7250 CCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCT 7309

Qy   2922 tgctgtttgcatcaaaaaaaaaaa 2945
          |||||||||||||||||||||||||
Db   7310 TGCTGTTTGCATCAAAAAAAAAAA 7333


RESULT  13
PEN133817
LOCUS       PEN133817    8849 bp    DNA             VRL       07-JAN-2000
DEFINITION  Porcine endogenous retrovirus type C proviral gag, pol and env
            genesand LTR (class A, clone 42).
ACCESSION   AJ133817
VERSION     AJ133817.1  GI:6688947
KEYWORDS    env gene; env protein; gag gene; gag protein; pol gene; pol
            protein.
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 8849)
  AUTHORS   Czauderna,F., Fischer,N., Boller,K., Krach,U., Kurth,R. and
            Toenjes,R.R.
  TITLE     Molecular Characterization of Human-tropic and
            Replication-competent Porcine Endogenous Retroviruses
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 8849)
  AUTHORS   Toenjes,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-1999) Toenjes R.R., Medical Biotechnology,
            Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, Hessen, D-63225
            Langen, GERMANY
FEATURES             Location/Qualifiers
     source          1. .8849
                     /organism="porcine endogenous retrovirus"
                     /proviral
                     /strain="type C"
                     /db_xref="taxon:61673"
                     /clone="42"
                     /note="grown in PK15 (ATCC No. CCL-33) cell line"
     LTR             1. .668
     misc_feature    506
                     /note="cap site"
     primer_bind     671. .688
                     /note="gly4"
     gene            1115. .2686
                     /gene="gag"
     mat_peptide     1115. .2686
                     /gene="gag"
                     /function="retroviral capsid"
                     /product="Gag protein"
     CDS             join(1115. .2686,2690. .6271,6150. .8132)
                     /note="viral polypeptide"
                     /codon_start=1
                     /protein_id="CAB65340.1"
                     /db_xref="GI:6688948"
                     /translation="MGQTVTTPLSLTLDHWTEVRSRAHNLSVQVKKGPWQTFCASEWP
                     TFDVGWPSEGTFNSEIILAVKAIIFQTGPGSHPDQEPYILTWQDLAEDPPPWVKPWLN
                     KPRKPGPRILALGEKNKHSAEKVEPSPRIYPEIEEPPTWPEPQPVPPPPYPAQGAVRG
                     PSAPPGAPVVEGPAAGTRSRRGATPERTDEIAILPLRTYGPPMPGGQLQPLQYWPFSS
                     ADLYNWKTNHPPFSEDPQRLTGLVESLMFSHQPTWDDCQQLLQTLFTTEERERILLEA
                     RKNVPGADGRPTQLQNEIDMGFPLTRPGWDYNTAEGRESLKIYRQALVAGLRGASRRP
                     TNLAKVREVMQGPNEPPSVFLERLMEAFRRFTPFDPTSEAQKASVALAFIGQSALDIR
```

```
                     KKLQRLEGLQEAELRDLVREAEKVYYRRETEEEKEQRKEKEREEREERRDRRQEKNLT
                     KILAAVVEGKSSRERERDFRKIRSGPRQSGNLGNRTPLDKDQCAYCKEKGHWARNCPK
                     KGNKGPKVLALEEDKDGRRGSDPLPEPRVTLKVEGQPVEFLVDTGAEHSVLLQPLGKL
                     KEKKSWVMGATGQRQYPWTTRRTVDLGVGRVTHSFLVIPECPVPLLGRDLLTKMGAQI
                     SFEQGRPEVSVNNKPITVLTLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMG
                     LAKQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPL
                     LPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFF
                     CLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFKIQH
                     PQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYS
                     LRGGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKE
                     KGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPW
                     RRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQ
                     PPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETG
                     VRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAE
                     LMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSL
                     LEALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPIIETPKAPEPR
                     RQYTLEDWQEIKKIDQFSETPEGTCYTSYGKEILPHKEGLEYVQQIHRLTHLGTKHLQ
                     QLVRTSPYHVLRLPGVADSVVKHCVPCQLVNANPSRIPPGKRLRGSHPGAHWEVDFTE
                     VKPAKYGNKYLLVFVDTFSGWVEAYPTKKETSTVVAKKILEEIFPRFGIPKVIGSDNG
                     PAFVAQVSQGLAKILGIDWKLHCAYRPQSSGQVERMNRTIKETLTKLTTETGINDWMA
                     LLPFVLFRVRNTPGQFGLTPYELLYGGPPPLVEIASVHSADVLLSQPLFSRLKALEWV
                     RQRAWKQLREAYSGEGDLQVPHRFQVGDSVYVRRHRAGNLETRWKGPYLVLLTTPTAV
                     KVERIPTWIHASHVKPAPPPDSGWKAEKTENPLKLRLHRVVPYSVNNSSSMHPTLSRR
                     HLPIRGGKPKRLKIPLSFASIAWFLTLSITPQVNGKRLVDSPNSHKPLSLTWLLTDSG
                     TGININSTQGEAPLGTWWPELYVCLRSVIPGLNDQATPPDVLRAYGFYVCPGPPNNEE
                     YCGNPQDFFCKQWSCITSNDGNWKWPVSQQDRVSYSFVNNPTSYNQFNYGHGRWKDWQ
                     QRVQKDVRNKQISCHSLDLDYLKISFTEKGKQENIQKWVNGISWGIVYYGGSGRKKGS
                     VLTIRLRIETQMEPPVAIGPNKGLAEQGPPIQEQRPSPNPSDYNTTSGSVPTEPNITI
                     KTGAKLFSLIQGAFQALNSTTPEATSSCWLCLASGPPYYEGMARGGKFNVTKEHRDQC
                     TWGSQNKLTLTEVSGKGTCIGMVPPSHQHLCNHTEAFNRTSESQYLVPGYDRWWACNT
                     GLTPCVSTLVFNQTKDFCVMVQIVPRVYYYPEKAVLDKYDYRYNRPKREPISLTLAVM
                     LGLGVAAGVGTGTAALITGPQQLEKGLSNLHRIVTEDLQALEKSVSNLEESLTSLSEV
                     VLQNRRGLDLLFLKEGGLCVALKEECCFYVDHSGAIRDSMSKLRERLERRRREREADQ
                     GWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPCLINRFVAFVRERVSAVQIMVLR
                     QQYQGLLSQGETDL"
     mat_peptide     2690. .6271
                     /gene="pol"
                     /function="retroviral protease and polymerase"
                     /product="Pol protein"
     gene            2690. .6271
                     /gene="pol"
     mat_peptide     6150. .8129
                     /gene="env"
                     /function="retroviral envelope"
                     /product="Env protein"
     gene            6150. .8129
                     /gene="env"
     LTR             8182. .8849
     polyA_signal    8742. .8747
BASE COUNT     2355 a   2191 c   2281 g   2022 t
ORIGIN
```

Query Match 21.2%; Score 626; DB 14; Length 8849;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 676; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   2082 gttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaa 2141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7877 GTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAA 7936

Qy   2142 caggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgct 2201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7937 CAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTGCT 7996

Qy   2202 cctgttacttacagttgggccttgcttaattaataggtttgttgccctttgttagagaacg 2261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7997 CCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACG 8056

Qy   2262 agtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagg 2321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8057 AGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGG 8116

Qy   2322 agaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtg 2381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8117 AGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTG 8176

Qy   2382 gggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaag 2441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8177 GGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAG 8236

Qy   2442 ctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtagaaggtcaca 2501
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   8237 CTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACA 8296

Qy   2502 cttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttga 2561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8297 CTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGA 8356

Qy   2562 ctaatcgcttatctggattctgtaaaactgactggcaccatagaagaattgattacacat 2621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8357 CTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACCATAGAAGAATTGATTACACAT 8416

Qy   2622 tgacagccctagtgacctatctcaactgcaatctgtcactctgcccaggagcccacgcag 2681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8417 TGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAG 8476

Qy   2682 atgcggacctccggagctattttaaaatgattggtccacggagcgcgggctctcgatatt 2741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8477 ATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATT 8536

Qy   2742 ttaaaatgattggtcca 2758
          |||||||||||||||||
Db   8537 TTAAAATGATTGGTCCA 8553
```

```
RESULT  14
AX002802
LOCUS       AX002802     2462 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 1 from Patent WO9853104.
ACCESSION   AX002802
VERSION     AX002802.1  GI:9885131
KEYWORDS    .
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 2462)
  AUTHORS   Stoye,J.P. and Weiss,R.A.
  TITLE     Detection of retroviral subtypes based upon envelope specific
            sequences
  JOURNAL   Patent: WO 9853104-A 1 26-NOV-1998;
            MEDICAL RES COUNCIL (GB); STOYE JONATHAN PAUL (GB); WEISS ROBIN
            ANTHONY (GB)
FEATURES             Location/Qualifiers
     source          1. .2462
                     /organism="porcine endogenous retrovirus"
                     /db_xref="taxon:61673"
BASE COUNT      715 a    573 c    575 g    599 t
ORIGIN
```

Query Match 15.5%; Score 457; DB 6; Length 2462;
Best Local Similarity 99.8%; Pred. No. 3.6e-252;
Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   2082 gttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaa 2141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1938 GTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAA 1997

Qy   2142 caggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgct 2201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1998 CAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTGCT 2057

Qy   2202 cctgttacttacagttgggccttgcttaattaataggtttgttgccctttgttagagaacg 2261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2058 CCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACG 2117

Qy   2262 agtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagg 2321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2118 AGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGG 2177

Qy   2322 agaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtg 2381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2178 AGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTG 2237

Qy   2382 gggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaag 2441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2238 GGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAG 2297

Qy   2442 ctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtagaaggtcaca 2501
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   2298 CTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACA 2357

Qy   2502 cttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttga 2561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2358 CTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGA 2417

Qy   2562 ctaatcgcttatctggattctgtaaaac 2589
          |||||||||||||||||||||||||||
Db   2418 CTAATCGCTTATCTGGATTCTGTAAAAC 2445


RESULT  15
PERENV1
LOCUS       PERENV1      2462 bp    RNA             VRL       14-OCT-1997
DEFINITION  Porcine endogenous retrovirus env gene, 2466 bp.
ACCESSION   Y12238
VERSION     Y12238.1  GI:2576322
KEYWORDS    env gene.
SOURCE      porcine endogenous retrovirus.
  ORGANISM  porcine endogenous retrovirus
            Viruses; Retroid viruses; Retroviridae; Mammalian type C
            retroviruses; 1-Mammalian type C virus group.
REFERENCE   1  (bases 1 to 2462)
  AUTHORS   Le Tissier,P., Stoye,J.P., Takeuchi,Y., Patience,C. and Weiss,R.A.
  TITLE     Two sets of human-tropic pig retrovirus
  JOURNAL   Nature 389 (6652), 681-682 (1997)
  MEDLINE   97478526
REFERENCE   2  (bases 1 to 2462)
  AUTHORS   Stoye,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-1997) J.P. Stoye, National Institute for Medical
            Research, Virology, The Ridgeway Mill Hill, London, NW7 1AA, UK
  REMARK    revised by author 16-SEP-97
FEATURES             Location/Qualifiers
     source          1. .2462
                     /organism="porcine endogenous retrovirus"
                     /specific_host="Sus scrofa"
                     /db_xref="taxon:61673"
                     /note="PERV-A"
     gene            211. .2193
                     /gene="env"
     CDS             211. .2193
                     /gene="env"
                     /codon_start=1
                     /protein_id="CAA72927.1"
                     /db_xref="GI:2576323"
                     /db_xref="SPTREMBL:O41172"
                     /translation="MHPTLSRRHLPIRGGKPKRLKIPLSFASIAWFLTLSITPQVNGK
                     RLVDSPNSHKPLSLTWLLTDSGTGININSTQGEAPLGTWWPELYVCLRSVIPGLNDQA
                     TPPDVLRAYGFYVCPGPPNNEEYCGNPQDFFCKQWSCITSNDGNWKWPVSQQDRVSYS
                     FVNNPTSYNQFNYGHGRWKDWQQRVQKDVRNKQISCHSLDLDYLKISFTEKGKQENIQ
                     KWVNGISWGIVYYGGSGRKKGSVLTIRLRIETQMEPPVAIGPNKGLAEQGPPIQEQRP
                     SPNPSDYNTTSGSVPTEPNITIKTGAKLFSLIQGAFQALNSTTPEATSSCWLCLASGP
                     PYYEGMARGGKFNVTKEHRDQCTWGSQNKLTLTEVSGKGTCIGMVPPSHQHLCNHTEA
                     FNRTSESQYLVPGYDRWWACNTGLTPCVSTLVFNQTKDFCVMVQIVPRVYYYPEKAVL
                     DEYDYRYNRPKREPISLTLAVMLGLGVAAGVGTGTAALITGPQQLEKGLSNLHRIVTE
                     DLQALEKSVSNLEESLTSLSEVVLQNRRGLDLLFLKEGGLCVALKEECCFYVDHSGAI
                     RDSMSKLRERLERRRREREADQGWFEGWFNRSPWMTTLLSALTGPLVVLLLLLTVGPC
                     LINRFVAFVRERVSAVQIMVLRQQYQGLLSQGETDL"
BASE COUNT      715 a    573 c    575 g    599 t
ORIGIN


  Query Match             15.5%;  Score 457;  DB 14;  Length 2462;
  Best Local Similarity   99.8%;  Pred. No. 3.6e-252;
  Matches  507;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   2082 gttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaa 2141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1938 GTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAA 1997

Qy   2142 caggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgct 2201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1998 CAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTGCT 2057

Qy   2202 cctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacg 2261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2058 CCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACG 2117

Qy   2262 agtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagg 2321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2118 AGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGG 2177

Qy   2322 agaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtg 2381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2178 AGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTG 2237

Qy   2382 gggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaag 2441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2238 GGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAG 2297

Qy   2442 ctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtagaaggtcaca 2501
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   2298 CTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACA 2357

Qy   2502 cttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttga 2561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2358 CTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGA 2417

Qy   2562 ctaatcgcttatctggattctgtaaaac 2589
          |||||||||||||||||||||||||||
Db   2418 CTAATCGCTTATCTGGATTCTGTAAAAC 2445


Search completed: February 24, 2002, 07:32:50
Job time: 28485 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:34:23 ; Search time 584.4 Seconds
(without alignments)
12042.750 Million cell updates/sec

Title: US-09-171-553B-3
Perfect score: 8209
Sequence: 1 gtggtgtacgactgtgggcc.....aaaaaaaaaaaaaaaaaaa 8209

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8209	100.0	8209	18	AAV09700	Porcine retrovirus
2	7802	95.0	8196	18	AAV09699	Porcine retrovirus
3	2967	36.1	7333	22	AAF77726	Defective retrovir
4	2967	36.1	7393	18	AAT74883	Porcine retrovirus
5	2682	32.7	3320	18	AAV09698	Porcine retrovirus
6	1499	18.3	1974	18	AAV09703	Porcine retrovirus
7	1466	17.9	3482	20	AAV82749	Pig endogenous ret
8	848	10.3	4402	22	AAC67020	PERV env protein c
9	797	9.7	6076	22	AAC67021	PERV env protein c
10	797	9.7	7362	22	AAC67019	PERV env protein c
11	746	9.1	6076	22	AAC67032	PERV env protein c
12	704	8.6	7892	18	AAT74884	Miniature swine re
13	704	8.6	8132	22	AAF77727	Nucleotide sequenc
14	570	6.9	4918	22	AAC67022	PERV env protein c
15	496	6.0	7873	22	AAC67023	PERV env protein c
16	457	5.6	2462	20	AAV82748	Pig endogenous ret
17	375	4.6	8060	18	AAT74811	Porcine retrovirus
18	375	4.6	8060	22	AAF77725	Tsukuba-1 cDNA. P
19	91	1.1	633	22	AAC67007	PERV-MSL LTR SEQ I
20	91	1.1	704	22	AAC67006	IamA1 LTR SEQ ID N
21	45	0.5	45	22	AAC82276	Human retrovirus D
22	41	0.5	45	22	AAC82275	Human retrovirus D
23	35	0.4	127	22	AAC86002	DoPEV RT sequence,
24	35	0.4	8655	20	AAV69750	Mus dunni endogeno
25	34	0.4	540	22	AAH26253	Pig endogenous ret
26	34	0.4	599	22	AAH26255	Pig endogenous ret
27	34	0.4	599	22	AAH26256	Pig endogenous ret
28	34	0.4	599	22	AAH26258	Pig endogenous ret
29	34	0.4	599	22	AAH26259	Pig endogenous ret
30	34	0.4	599	22	AAH26260	Pig endogenous ret
31	34	0.4	600	22	AAH26251	Pig endogenous ret
32	34	0.4	600	22	AAH26252	Pig endogenous ret
33	34	0.4	600	22	AAH26254	Pig endogenous ret
34	34	0.4	600	22	AAH26257	Pig endogenous ret
35	34	0.4	1980	22	AAC67018	PERV-1-15 env prot
36	31	0.4	2702	19	AAV55872	Plant acetolactate
37	30	0.4	1431	21	AAC77334	Human ORFX ORF2889
38	29	0.4	1234	18	AAT94005	Retrovirus 4070A p
39	29	0.4	3557	7	AAN60801	Human pro-growth h
40	29	0.4	3643	21	AAA87841	MLV reverse transc
41	29	0.4	5178	21	AAA15660	Gibbon leukaemia v
42	29	0.4	7308	18	AAT90693	Plasmid hCMV+intro
43	29	0.4	7308	18	AAT90694	Plasmid hCMV+intro
44	29	0.4	7616	18	AAT90688	Plasmid CeB coding
c 45	29	0.4	8202	17	AAT09280	Novel AMP/MCF viru

ALIGNMENTS

RESULT 1
AAV09700
ID AAV09700 standard; DNA; 8209 BP.
XX
AC AAV09700;
XX
DT 20-MAY-1998 (first entry)
XX
DE Porcine retrovirus DNA encoding the LTR and GAG, POL and ENV proteins.
XX
KW Porcine retrovirus; PoEV; POL protein; ENV protein; GAG protein;
KW virion core polypeptide; polymerase protein; envelope protein;
KW vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
XX
OS Porcine retrovirus.
XX
FH Key Location/Qualifiers
FT misc_feature 1..61
FT /*tag= a
FT /note= "R LTR domain"
FT LTR 1..588
FT /*tag= b
FT misc_feature 62..143
FT /*tag= c
FT /note= "U5 LTR domain"
FT CDS 588..2162
FT /*tag= d
FT /product= GAG protein
FT /note= "viral core polypeptide"
FT CDS 2163..5747
FT /*tag= e
FT /product= POL protein
FT /note= "polymerase peptide"
FT CDS 5620..7590

```
FT                   /*tag= f
FT                   /product= ENV protein
FT                   /note= "envelope protein"
FT   LTR             7591..8209
FT                   /*tag= g
FT   misc_feature    7638..8106
FT                   /*tag= h
FT                   /note= "U3 LTR domain"
FT   protein_bind    7643..7686
FT                   /*tag= i
FT                   /note= "Binding site for transcription factors
FT                        OCT-1, c-Myb, LyF-1, E47 and ETS-1"
FT   protein_bind    7719..7739
FT                   /*tag= j
FT                   /note= "Binding site for transcription factors
FT                        AP-4 and NF1"
FT   protein_bind    7744..7754
FT                   /*tag= k
FT                   /note= "Binding site for transcription factor AP-1/TR"
FT   protein_bind    7773..7785
FT                   /*tag= l
FT                   /note= "Binding site for transcription factor ETS-1/GATA"
FT   protein_bind    7793..7834
FT                   /*tag= m
FT                   /note= "Binding site for transcription factors
FT                        GATA, ETS-1, c-Myb, AP-1, GATA, E47"
FT   protein_bind    7841..7851
FT                   /*tag= n
FT                   /note= "Binding site for transcription factor AP-1"
FT   protein_bind    7873..7883
FT                   /*tag= o
FT                   /note= "Binding site for transcription factor AP-1"
FT   protein_bind    7885..7914
FT                   /*tag= p
FT                   /note= "Binding site for transcription factors
FT                        AP-1/GATA, c-Myb and AP-1"
FT   protein_bind    7928..7953
FT                   /*tag= q
FT                   /note= "Binding site for transcription factors E47
FT                        and ETS-1"
FT   CAAT_signal     7964..7968
FT                   /*tag= r
FT   protein_bind    7985..7997
FT                   /*tag= s
FT                   /note= "Binding site for transcription factor GATA"
FT   CAAT_signal     8003..8007
FT                   /*tag= t
FT   protein_bind    8024..8036
FT                   /*tag= u
FT                   /note= "Binding site for transcription factor GATA"
FT   CAAT_signal     8042..8046
FT                   /*tag= v
FT   protein_bind    8049..8060
FT                   /*tag= w
FT                   /note= "Binding site for transcription factor
FT                        AP-1/CREB"
FT   TATA_signal     8075..8090
FT                   /*tag= x
FT   misc_feature    1..61
FT                   /*tag= y
FT                   /note= "R LTR domain"
FT   misc_feature    8127..8135
FT                   /*tag= z
FT                   /note= "PolyA downstream element"
FT   polyA_signal    8161..8166
FT                   /*tag= aa
XX
PN   WO9740167-A1.
XX
PD   30-OCT-1997.
XX
PF   18-APR-1997;    97WO-GB01087.
XX
PR   10-FEB-1997;    97GB-0002668.
PR   19-APR-1996;    96GB-0008164.
XX
PA   (IMUT-) IMUTRAN LTD.
PA   (QONE-) Q-ONE BIOTECH LTD.
XX
PI   Galbraith DN,  Haworth C,  Lees GM,  Smith KT;
XX
DR   WPI; 1997-535851/49.
DR   P-PSDB; AAW39271; AAW39272; AAW39273.
XX
PT   Polynucleotide encoding porcine retrovirus expression product -
PT   useful to develop products for use in vaccines, diagnosis and
PT   xeno-transplantation
XX
PS   Claim 4; Fig 3; 69pp; English.
XX
CC   This DNA sequence encodes the porcine retrovirus (PoEV) virion core
CC   polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins and
CC   also includes the Long Terminal Repeat (LTR). These proteins can be used
CC   to develop viral vaccines, antisense nucleic acids, ribozymes and other
CC   antiviral agents. They can also be used in xeno-transplantation
CC   technology and as diagnostic tools.
XX
SQ   Sequence 8209 BP; 2168 A; 2064 C; 2154 G; 1823 T; 0 other;


  Query Match             100.0%;  Score 8209;  DB 18;  Length 8209;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 8209;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60

Qy      61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggat 120

Qy     121 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctta 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctta 180

Qy     181 cacccgagaaccgacttggaggtaaagggatccccctttggaacgtgtgtgtgtgtcggcc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 cacccgagaaccgacttggaggtaaagggatccccctttggaacgtgtgtgtgtgtcggcc 240

Qy     241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300

Qy     301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360

Qy     361 ccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 420

Qy     421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttccccctccgcggccgtccg 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttccccctccgcggccgtccg 480

Qy     481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540

Qy     541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600

Qy     601 tgactacccccctagtttgactctcgaccattggactgaagttagatccagggctcata 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    601 tgactacccccttagtttgactctcgaccattggactgaagttagatccagggctcata 660
Qy    661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
Db    661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
Qy    721 cattcgatgttggatggccatcagaggggacctttaattctgaaattatcctggctgtta 780
Db    721 cattcgatgttggatggccatcagaggggacctttaattctgaaattatcctggctgtta 780
Qy    781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 840
Db    781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 840
Qy    841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
Db    841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
Qy    901 gaaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
Db    901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
Qy    961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
Db    961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
Qy   1021 caacctgttcccccaccccctatccagcacagggtgctgtgaggggacctctgcccctc 1080
Db   1021 caacctgttcccccaccccctatccagcacagggtgctgtgaggggacctctgcccctc 1080
Qy   1081 ctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccaccc 1140
Db   1081 ctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccaccc 1140
Qy   1141 cggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgccag 1200
Db   1141 cggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgccag 1200
Qy   1201 ggggccaattgcagccctccagtattggccctttttcttctgcagatctctataattgga 1260
Db   1201 ggggccaattgcagccctccagtattggccctttttcttctgcagatctctataattgga 1260
Qy   1261 aaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtccc 1320
Db   1261 aaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtccc 1320
Qy   1321 ttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttca 1380
Db   1321 ttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttca 1380
Qy   1381 caaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgacg 1440
Db   1381 caaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgacg 1440
Qy   1441 ggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggtt 1500
Db   1441 ggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggtt 1500
Qy   1501 gggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtgg 1560
Db   1501 gggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtgg 1560
Qy   1561 cgggtctccgggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgc 1620
Db   1561 cgggtctccgggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgc 1620
Qy   1621 agggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcggt 1680
Db   1621 agggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcggt 1680
Qy   1681 tcacccctttgatcctacctcagaggcccagaaagcctcagtggccctggccttcattg 1740
Db   1681 tcacccctttgatcctacctcagaggcccagaaagcctcagtggccctggccttcattg 1740
Qy   1741 ggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggctg 1800
Db   1741 ggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggctg 1800
Qy   1801 agttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagagg 1860
Db   1801 agttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagagg 1860
Qy   1861 agaaggaacagagaaaagaaaaggagagagaagaaaagggaggaaagacgtgatagacggc 1920
Db   1861 agaaggaacagagaaaagaaaaggagagagaagaaaagggaggaaagacgtgatagacggc 1920
Qy   1921 aagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggaga 1980
Db   1921 aagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggaga 1980
Qy   1981 gagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaata 2040
Db   1981 gagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaata 2040
Qy   2041 ggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagga 2100
Db   2041 ggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagga 2100
Qy   2101 actgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagatt 2160
Db   2101 actgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagatt 2160
Qy   2161 aggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtggaggggc 2220
Db   2161 aggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtggaggggc 2220
Qy   2221 aaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccattag 2280
Db   2221 aaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccattag 2280
Qy   2281 gaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatccat 2340
Db   2281 gaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatccat 2340
Qy   2341 ggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtca 2400
Db   2341 ggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtca 2400
Qy   2401 tccctgagtgcccagtacccttctaggtagagacttactgaccaagatgggagctcaaa 2460
Db   2401 tccctgagtgcccagtacccttctaggtagagacttactgaccaagatgggagctcaaa 2460
Qy   2461 tttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttga 2520
Db   2461 tttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttga 2520
Qy   2521 ccctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagata 2580
Db   2521 ccctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagata 2580
Qy   2581 tacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttgg 2640
Db   2581 tacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttgg 2640
Qy   2641 caaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtca 2700
Db   2641 caaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtca 2700
Qy   2701 gacagtacccccttgagtagagaggctcgagaaggaatttggccgcatgttcaaagattaa 2760
Db   2701 gacagtacccccttgagtagagaggctcgagaaggaatttggccgcatgttcaaagattaa 2760
Qy   2761 tccaacagggcatcctagttcctgtccaatccccttggaatactcccctgctaccggtta 2820
Db   2761 tccaacagggcatcctagttcctgtccaatccccttggaatactcccctgctaccggtta 2820
```

```
Qy   2821 ggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaaggg 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 ggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaaggg 2880

Qy   2881 tgcaggacatacacccaacggtcccgaaccccttataacctcttgagcgccctcccgcctg 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 tgcaggacatacacccaacggtcccgaaccccttataacctcttgagcgccctcccgcctg 2940

Qy   2941 aacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacacc 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 aacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacacc 3000

Qy   3001 ccactagccaaccacttttgccttcgaatggagagatccaggtacgggaagaaccgggc 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 ccactagccaaccacttttgccttcgaatggagagatccaggtacgggaagaaccgggc 3060

Qy   3061 agctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaag 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 agctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaag 3120

Qy   3121 ccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccagt 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 ccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccagt 3180

Qy   3181 acgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaagg 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 acgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaagg 3240

Qy   3241 cactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagattt 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 cactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagattt 3300

Qy   3301 gcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacgg 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 gcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacgg 3360

Qy   3361 aggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgagag 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 aggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgagag 3420

Qy   3421 agtttttggggacagctggattttgcagactgtggatcccggggtttgcgaccttagcag 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 agtttttggggacagctggattttgcagactgtggatcccggggtttgcgaccttagcag 3480

Qy   3481 ccccactctacccgctaaccaaagaaaaaggggattctcctgggctcctgagcaccaga 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 ccccactctacccgctaaccaaagaaaaaggggattctcctgggctcctgagcaccaga 3540

Qy   3541 aggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctccctgacg 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 aggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctccctgacg 3600

Qy   3601 taactaaaccctttaccctttatgtggatgagcgtaagggagtagcccgaggagttttaa 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 taactaaaccctttaccctttatgtggatgagcgtaagggagtagcccgaggagttttaa 3660

Qy   3661 cccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagcttgatcctg 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 cccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagcttgatcctg 3720

Qy   3721 tagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccatactggtcaagg 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 tagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccatactggtcaagg 3780

Qy   3781 acgctgacaaattgactttgggacagaatataactgtaatagcccccatgcattggaga 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 acgctgacaaattgactttgggacagaatataactgtaatagcccccatgcattggaga 3840

Qy   3841 acatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacccactatcaaa 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 acatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacccactatcaaa 3900

Qy   3901 gcctgcttctcacagagagggtcactttcgctccaccagccgctctcaaccctgccactc 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 gcctgcttctcacagagagggtcactttcgctccaccagccgctctcaaccctgccactc 3960

Qy   3961 ttctgcctgaagagactgatgaaccagtgactcatgattgccatcaactattgattgagg 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 ttctgcctgaagagactgatgaaccagtgactcatgattgccatcaactattgattgagg 4020

Qy   4021 agactggggtccgcaaggaccttacagacataccgctgactggagaagtgctaacctggt 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 agactggggtccgcaaggaccttacagacataccgctgactggagaagtgctaacctggt 4080

Qy   4081 tcactgacggaagcagctatgtggtggaaggtaagaggatggctggggcggcagtggtgg 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 tcactgacggaagcagctatgtggtggaaggtaagaggatggctggggcggcagtggtgg 4140

Qy   4141 acgggacccgcacgatctgggccagcagcctgccggaaggaacttcagcgcaaaaggctg 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 acgggacccgcacgatctgggccagcagcctgccggaaggaacttcagcgcaaaaggctg 4200

Qy   4201 agctcatggccctcacgcaagctttgcggctggccgaagggaaatccataaacatttata 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 agctcatggccctcacgcaagctttgcggctggccgaagggaaatccataaacatttata 4260

Qy   4261 cggacagcaggtatgcctttgcgactgcacacgtacacggggccatctataaacaaaggg 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 cggacagcaggtatgcctttgcgactgcacacgtacacggggccatctataaacaaaggg 4320

Qy   4321 ggttgcttacctcagcagggagggaaataaagaacaaagaggaaattctaagcctattag 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 ggttgcttacctcagcagggagggaaataaagaacaaagaggaaattctaagcctattag 4380

Qy   4381 aagccttacatttgccaaaaaggctagctattatacactgtcctggacatcagaaagcca 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 aagccttacatttgccaaaaaggctagctattatacactgtcctggacatcagaaagcca 4440

Qy   4441 aagatctcatatctagagggaaccagatggctgaccgggttgccaagcaggcagcccagg 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 aagatctcatatctagagggaaccagatggctgaccgggttgccaagcaggcagcccagg 4500

Qy   4501 ctgttaaccttctgcctataatagaaacgcccaaagccccagaacccagacgacagtaca 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 ctgttaaccttctgcctataatagaaacgcccaaagccccagaacccagacgacagtaca 4560

Qy   4561 ccctagaagactggcaagagataaaaaagatagaccagttctctgagactccggagggga 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 ccctagaagactggcaagagataaaaaagatagaccagttctctgagactccggagggga 4620

Qy   4621 cctgctatacctcatatgggaaggaaatcctgccccacaaagaagggttagaatatgtcc 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 cctgctatacctcatatgggaaggaaatcctgccccacaaagaagggttagaatatgtcc 4680

Qy   4681 aacagatacatcgtctaacccacctaggaactaaacacctgcagcagttggtcagaacat 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 aacagatacatcgtctaacccacctaggaactaaacacctgcagcagttggtcagaacat 4740

Qy   4741 ccccttatcatgttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgc 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 ccccttatcatgttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgc 4800

Qy   4801 cctgccagctggttaatgctaatccttccagaatacctccaggaaagagactaaggggaa 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 cctgccagctggttaatgctaatccttccagaatacctccaggaaagagactaaggggaa 4860

Qy   4861 gccacccaggcgctcactgggaagtggacttcactgaggtaaagccggctaaatacggaa 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 gccacccaggcgctcactgggaagtggacttcactgaggtaaagccggctaaatacggaa 4920

Qy   4921 acaaatatctattggttttttgtagacacctttttcaggatgggtagaggcttatcctacta 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 acaaatatctattggttttttgtagacacctttttcaggatgggtagaggcttatcctacta 4980

Qy   4981 agaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttccaagatttg 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   4981 agaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttccaagatttg 5040

Qy   5041 gaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcagg 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 gaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcagg 5100

Qy   5101 gactggccaagatattggggattgattggaaactgcattgtgcatacagaccccaaagct 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 gactggccaagatattggggattgattggaaactgcattgtgcatacagaccccaaagct 5160

Qy   5161 caggacaggtagagaggatgaatagaaccattaaagagacccttaccaaattgaccacag 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 caggacaggtagagaggatgaatagaaccattaaagagacccttaccaaattgaccacag 5220

Qy   5221 agactggcattaatgattggatggctctcctgccctttgtgctttttagggtgaggaaca 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 agactggcattaatgattggatggctctcctgccctttgtgctttttagggtgaggaaca 5280

Qy   5281 cccctggacagtttgggctgacccccctatgaattgctctacgggggaccccccccgttgg 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 cccctggacagtttgggctgacccccctatgaattgctctacgggggaccccccccgttgg 5340

Qy   5341 cagaaattgcctttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggc 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 cagaaattgcctttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggc 5400

Qy   5401 tcaaggcgctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcag 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 tcaaggcgctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcag 5460

Qy   5461 gaggagacttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgcc 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 gaggagacttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgcc 5520

Qy   5521 accgtgcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacac 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 accgtgcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacac 5580

Qy   5581 caacggctgtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccgg 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 caacggctgtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccgg 5640

Qy   5641 cgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcgcc 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 cgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcgcc 5700

Qy   5701 tccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttata 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701 tccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttata 5760

Qy   5761 gacagctcgaaccccatagaccttatccttacctggctgattattgaccctgatacg 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 gacagctcgaaccccatagaccttatccttacctggctgattattgaccctgatacg 5820

Qy   5821 ggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaactg 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 ggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaactg 5880

Qy   5881 catttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtccgt 5940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5881 catttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtccgt 5940

Qy   5941 agttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctggg 6000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5941 agttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctggg 6000

Qy   6001 gaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggccg 6060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6001 gaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggccg 6060

Qy   6061 atctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtacaaa 6120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6061 atctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtacaaa 6120

Qy   6121 atgatgaaactatataaagataagagctgctccccatcagacttagattatctaaagata 6180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6121 atgatgaaactatataaagataagagctgctccccatcagacttagattatctaaagata 6180

Qy   6181 agtttcactgaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagctgggg 6240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6181 agtttcactgaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagctgggg 6240

Qy   6241 aatagtttttttattatatggcgggggagcagggtccactttaaccattcgccttaggata 6300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6241 aatagtttttttattatatggcgggggagcagggtccactttaaccattcgccttaggata 6300

Qy   6301 gagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaacagggg 6360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6301 gagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaacagggg 6360

Qy   6361 cccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgcggcct 6420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6361 cccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgcggcct 6420

Qy   6421 gacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgcctaga 6480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6421 gacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgcctaga 6480

Qy   6481 aactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagggagct 6540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6481 aactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagggagct 6540

Qy   6541 ttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtctatcc 6600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6541 ttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtctatcc 6600

Qy   6601 tcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagag 6660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6601 tcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagag 6660

Qy   6661 catagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagtttccggg 6720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6661 catagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagtttccggg 6720

Qy   6721 aaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagtactgtg 6780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6721 aaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagtactgtg 6780

Qy   6781 gtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggca 6840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6781 gtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggca 6840

Qy   6841 tgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttc 6900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6841 tgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttc 6900

Qy   6901 tgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgat 6960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6901 tgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgat 6960

Qy   6961 gaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctagctgta 7020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6961 gaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctagctgta 7020

Qy   7021 atgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatcacagga 7080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7021 atgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatcacagga 7080

Qy   7081 ccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatctccga 7140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7081 ccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatctccga 7140

Qy   7141 gccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaagtggtt 7200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7141 gccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaagtggtt 7200
```

```
Qy    7201 ctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtgcagcc 7260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7201 ctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtgcagcc 7260

Qy    7261 ttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactccatgaac 7320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7261 ttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactccatgaac 7320

Qy    7321 aagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggtggttt 7380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7321 aagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggtggttt 7380

Qy    7381 gaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggccc 7440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7381 gaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggccc 7440

Qy    7441 ctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgcc 7500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7441 ctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgcc 7500

Qy    7501 tttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggc 7560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7501 tttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggc 7560

Qy    7561 cttctgagccaaggagaaactgacctctagccttcccagttctaagattagaactattaa 7620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7561 cttctgagccaaggagaaactgacctctagccttcccagttctaagattagaactattaa 7620

Qy    7621 caagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccagg 7680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7621 caagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccagg 7680

Qy    7681 aagttaataaaaagctctaaatgcccccgaattacagaccctgctggctgccagtaaata 7740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7681 aagttaataaaaagctctaaatgcccccgaattacagaccctgctggctgccagtaaata 7740

Qy    7741 ggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaagataac 7800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7741 ggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaagataac 7800

Qy    7801 aggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccatagaag 7860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7801 aggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccatagaag 7860

Qy    7861 aattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactctgccc 7920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7861 aattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactctgccc 7920

Qy    7921 aggagcccacgcagatgcggacctccggagctattttaaaatgattggtccacggagcgc 7980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7921 aggagcccacgcagatgcggacctccggagctattttaaaatgattggtccacggagcgc 7980

Qy    7981 gggctctcgatattttaaaatgattggtccatggagcgcgggctctcgatattttaaaat 8040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7981 gggctctcgatattttaaaatgattggtccatggagcgcgggctctcgatattttaaaat 8040

Qy    8041 gattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgattccgca 8100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8041 gattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgattccgca 8100

Qy    8101 ctcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgcttgg 8160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8101 ctcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgcttgg 8160

Qy    8161 aataaaaatcctcttgctgtttgcatcaaaaaaaaaaaaaaaaaaaaaa 8209
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8161 aataaaaatcctcttgctgtttgcatcaaaaaaaaaaaaaaaaaaaaaa 8209
```

RESULT 2
AAV09699
ID AAV09699 standard; DNA; 8196 BP.
XX

AC AAV09699;
XX
DT 19-MAY-1998 (first entry)
XX
DE Porcine retrovirus DNA encoding, GAG, POL and ENV.
XX
KW Porcine retrovirus; PoEV; POL protein; ENV protein; GAG protein;
KW vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
XX
OS Porcine retrovirus.
XX
FH Key Location/Qualifiers
FT CDS 576..2126
FT /*tag= a
FT /product= GAG protein
FT /note= "virion core polypeptide"
FT CDS 2143..5733
FT /*tag= b
FT /product= POL protein
FT /note= "polymerase peptide sequence as given in
FT specification"
FT CDS 5606..7576
FT /*tag= c
FT /product= ENV protein
FT /note= "envelope protein"
XX
PN WO9740167-A1.
XX
PD 30-OCT-1997.
XX
PF 18-APR-1997; 97WO-GB01087.
XX
PR 10-FEB-1997; 97GB-0002668.
PR 19-APR-1996; 96GB-0008164.
XX
PA (IMUT-) IMUTRAN LTD.
PA (QONE-) Q-ONE BIOTECH LTD.
XX
PI Galbraith DN, Haworth C, Lees GM, Smith KT;
XX
DR WPI; 1997-535851/49.
XX
PT Polynucleotide encoding porcine retrovirus expression product -
PT useful to develop products for use in vaccines, diagnosis and
PT xeno-transplantation
XX
PS Claim 4; Fig 2; 69pp; English.
XX
CC This DNA sequence encodes the porcine retrovirus (PoEV) virion core
CC polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins. These
CC proteins can be used to develop viral vaccines, antisense nucleic acids,
CC ribozymes and other antiviral agents. They can also be used in
CC xeno-transplantation technology and as diagnostic tools.
XX
SQ Sequence 8196 BP; 2165 A; 2061 C; 2147 G; 1820 T; 3 other;

Query Match 95.0%; Score 7802; DB 18; Length 8196;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 8192; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```
Qy      13 tgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatcaagaccgcttct 72
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 tgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatcaagaccgcttct 60

Qy      73 cgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggattgttcttttact 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 cgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggattgttcttttact 120

Qy     133 ggcctttcatttggtgcgttggccgggaaatcctgcgaccacccccttacacccgagaacc 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ggcctttcatttggtgcgttggccgggaaatcctgcgaccacccccttacacccgagaacc 180
```

```
Qy    193 gacttggaggtaaagggatccccttggaacgtgtgtgtgtgtcggccggcgtctctgtt 252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 gacttggaggtaaagggatccccttggaacgtgtgtgtgtcggccggcgtctctgtt 240

Qy    253 ctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcctctcagaccgtaa 312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcctctcagaccgtaa 300

Qy    313 ggactggaggactgtgatcagcagacgtgctaggaggatcacaggctgccaccctggggg 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ggactggaggactgtgatcagcagacgtgctaggaggatcacaggctgccaccctggggg 360

Qy    373 acgccccgggaggtggggagagccagggacgcctggtggtctcctactgtcggtcagagg 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 acgccccgggaggtggggagagccagggacgcctggtggtctcctactgtcggtcagagg 420

Qy    433 accgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccgactcttttgcct 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 accgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccgactcttttgcct 480

Qy    493 gcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggtttctgtctcgtgtg 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 gcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggtttctgtctcgtgtg 540

Qy    553 tctttgtcttgtgcgtccttgtctacagttttaatatgggacagacagtgactacccccc 612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 tctttgtcttgtgcgtccttgtctacagttttaatatgggacagacagtgactacccccc 600

Qy    613 ttagtttgactctcgaccattggactgaagttagatccagggctcataatttgtcagttc 672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ttagtttgactctcgaccattggactgaagttagatccagggctcataatttgtcagttc 660

Qy    673 aggttaagaagggaccttggcagactttctgtgcctctgaatggccaacattcgatgttg 732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 aggttaagaagggaccttggcagactttctgtgcctctgaatggccaacattcgatgttg 720

Qy    733 gatggccatcagaggggacctttaattctgaaattatcctggctgttaaggcaatcattt 792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gatggccatcagaggggacctttaattctgaaattatcctggctgttaaggcaatcattt 780

Qy    793 ttcagactggacccggctctcatcctgatcaggagccctatatccttacgtggcaagatt 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ttcagactggacccggctctcatcctgatcaggagccctatatccttacgtggcaagatt 840

Qy    853 tggcagaagatcctccgccatgggttaaaccatggctaaataaaccaagaaagccaggtc 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 tggcagaagatcctccgccatgggttaaaccatggctaaataaaccaagaaagccaggtc 900

Qy    913 cccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaagtcgagccctctt 972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 cccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaagtcgagccctctt 960

Qy    973 cctcgtatctaccccgagatcgaggagccgccgacttggccggaaccccaacctgttccc 1032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 cctcgtatctaccccgagatcgaggagccgccgacttggccggaaccccaacctgttccc 1020

Qy   1033 ccaccccctatccagcacagggtgctgtgaggggacctctgcccctcctggagctccgg 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ccaccccctatccagcacagggtgctgtgaggggacctctgcccctcctggagctccgg 1080

Qy   1093 tggtggagggacctgctgccgggactcggagccggagaggcgccaccccggagcggacag 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 tggtggagggacctgctgccgggactcggagccggagaggcgccaccccggagcggacag 1140

Qy   1153 acgagatcgcgatattaccgctgcgcacctatggccctcccatgccaggggccaattgc 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 acgagatcgcgatattaccgctgcgcacctatggccctcccatgccaggggccaattgc 1200

Qy   1213 agcccctccagtattggccctttttcttctgcagatctctataattggaaaactaaccatc 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 agcccctccagtattggccctttttcttctgcagatctctataattggaaaactaaccatc 1260

Qy   1273 ccccttttctcggaggatccccaacgcctcacggggttggtggagtcccttatgttctctc 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ccccttttctcggaggatccccaacgcctcacggggttggtggagtcccttatgttctctc 1320

Qy   1333 accagcctacttgggatgattgtcaacagctgctgcagacactcttcacaaccgaggagc 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 accagcctacttgggatgattgtcaacagctgctgcagacactcttcacaaccgaggagc 1380

Qy   1393 gagagagaattctgttagaggctagaaaaaaatgttcctggggccgacgggcgacccacgc 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 gagagagaattctgttagaggctagaaaaaaatgttcctggggccgacgggcgacccacgc 1440

Qy   1453 agttgcaaaatgagattgacatgggatttcccttgactcgccccggttgggactacaaca 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 agttgcaaaatgagattgacatgggatttcccttgactcgccccggttgggactacaaca 1500

Qy   1513 cggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcgggtctccggg 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 cggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcgggtctccggg 1560

Qy   1573 gcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcagggaccgaacg 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 gcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcagggaccgaacg 1620

Qy   1633 aacctccctcggtatttcttgagaggctcatggaagccttcaggcggttcacccctttg 1692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 aacctccctcggtatttcttgagaggctcatggaagccttcaggcggttcacccctttg 1680

Qy   1693 atcctacctcagaggcccagaaagcctcagtggccctggccttcattgggcagtcggctc 1752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 atcctacctcagaggcccagaaagcctcagtggccctggccttcattgggcagtcggctc 1740

Qy   1753 tggatatcaggaagaaacttcagagactggaagggttacaggaggctgagttacgtgatc 1812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 tggatatcaggaagaaacttcagagactggaagggttacaggaggctgagttacgtgatc 1800

Qy   1813 tagtgagagaggcagagaaggtgtattacagaagggagacagaagaggagaaggaacaga 1872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 tagtgagagaggcagagaaggtgtattacagaagggagacagaagaggagaaggaacaga 1860

Qy   1873 gaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacggcaagagaagaatt 1932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 gaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacggcaagagaagaatt 1920

Qy   1933 tgactaagatcttggccgcagtggttgaagggaagagcagcagggagagagagagagatt 1992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 tgactaagatcttggccgcagtggttgaagggaagagcagcagggagagagagagagatt 1980

Qy   1993 ttaggaaaattaggtcaggccctagacagtcagggaacctgggcaataggaccccactcg 2052
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 ttaggaaaattaggtcaggccctagacagtcagggaacctgggcaataggaccccactcg 2040

Qy   2053 acaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaaggaactgccccaaga 2112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 acaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaaggaactgccccaaga 2100

Qy   2113 agggaaacaaaggaccgaaggtcctagctctagaagaagatataagattaggggagacggg 2172
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||| ||
Db   2101 agggaaacaaaggaccgaa-gtcctagctctagaagaagatataagattaggggagac-gg 2158

Qy   2173 gttcggacccccctccccgagcccagggtaactttgaaggtggaggggcaaccagttgagt 2232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2159 gttcggacccccctccccgagcccagggtaactttgaaggtggaggggcaaccagttgagt 2218

Qy   2233 tcctggttgataccggagcggagcattcagtgctgctacaaccattaggaaaactaaaag 2292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2219 tcctggttgataccggagcggagcattcagtgctgctacaaccattaggaaaactaaaag 2278

Qy   2293 aaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatccatggactacccgaa 2352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2279 aaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatccatggactacccgaa 2338

Qy   2353 gaaccgttgacttgggagtgggacgggtaacccactcgtttctggtcatccctgagtgcc 2412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  2339 gaaccgttgacttgggagtgggacgggtaacccactcgtttctggtcatccctgagtgcc 2398

Qy  2413 cagtacccctttctaggtagagacttactgaccaagatgggagctcaaatttcttttgaac 2472
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2399 cagtacccctttctaggtagagacttactgaccaagatgggagctcaaatttcttttgaac 2458

Qy  2473 aaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttgaccctccaattag 2532
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2459 aaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttgaccctccaattag 2518

Qy  2533 atgatgaatatcgactatattctccccaagtaaagcctgatcaagatatacagtcctggt 2592
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2519 atgatgaatatcgactatattctccccaagtaaagcctgatcaagatatacagtcctggt 2578

Qy  2593 tggagcagtttccccaagcctgggcagaaaccgcagggatgggtttggcaaagcaagttc 2652
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2579 tggagcagtttccccaagcctgggcagaaaccgcagggatgggtttggcaaagcaagttc 2638

Qy  2653 ccccacaggttattcaactgaaggccagtgctacaccagtatcagtcagacagtacccct 2712
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2639 ccccacaggttattcaactgaaggccagtgctacaccagtatcagtcagacagtacccct 2698

Qy  2713 tgagtagagaggctcgagaaggaatttggccgcatgttcaaagattaatccaacagggca 2772
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2699 tgagtagagaggctcgagaaggaatttggccgcatgttcaaagattaatccaacagggca 2758

Qy  2773 tcctagttcctgtccaatccccttggaatactcccctgctaccggttaggaagcctggga 2832
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2759 tcctagttcctgtccaatccccttggaatactccctgctaccggttaggaagcctggga 2818

Qy  2833 ccaatgattatcgaccagtacaggacttgagagaggtcaataaaagggtgcaggacatac 2892
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2819 ccaatgattatcgaccagtacaggacttgagagaggtcaataaaagggtgcaggacatac 2878

Qy  2893 acccaacggtcccgaaccctataacctcttgagcgccctcccgcctgaacggaactggt 2952
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2879 acccaacggtcccgaaccctataacctcttgagcgccctcccgcctgaacggaactggt 2938

Qy  2953 acacagtattggacttaaaagatgccttcttctgcctgagattacaccccactagccaac 3012
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2939 acacagtattggacttaaaagatgccttcttctgcctgagattacaccccactagccaac 2998

Qy  3013 cacttttgccttcgaatggagagatccaggtacgggaagaaccgggcagctcacctgga 3072
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2999 cacttttgccttcgaatggagagatccaggtacgggaagaaccgggcagctcacctgga 3058

Qy  3073 cccgactgccccaagggttcaagaactccccgaccatctttgacgaagccctacacaggg 3132
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3059 cccgactgccccaagggttcaagaactccccgaccatctttgacgaagccctacacaggg 3118

Qy  3133 acctggccaacttcaggatccaacaccctcaggtgaccctcctccagtacgtggatgacc 3192
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3119 acctggccaacttcaggatccaacaccctcaggtgaccctcctccagtacgtggatgacc 3178

Qy  3193 tgcttctggcgggagccaccaaacaggactgcttagaaggtacgaaggcactactgctgg 3252
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3179 tgcttctggcgggagccaccaaacaggactgcttagaaggtacgaaggcactactgctgg 3238

Qy  3253 aattgtctgacctaggctacagagcctctgctaagaaggcccagatttgcaggagagagg 3312
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3239 aattgtctgacctaggctacagagcctctgctaagaaggcccagatttgcaggagagagg 3298

Qy  3313 taacatacttggggtacagtttgcggggcgggcagcgatggctgacggaggcacggaaga 3372
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3299 taacatacttggggtacagtttgcggggcgggcagcgatggctgacggaggcacggaaga 3358

Qy  3373 aaactgtagtccagataccggccccaaccacagccaaacaagtgagagagtttttgtggga 3432
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3359 aaactgtagtccagataccggccccaaccacagccaaacaagtgagagagtttttgtggga 3418

Qy  3433 cagctggattttgcagactgtggatcccggggtttgcgaccttagcagccccactctacc 3492
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3419 cagctggattttgcagactgtggatcccggggtttgcgaccttagcagccccactctacc 3478

Qy  3493 cgctaaccaaagaaaaaggggattctcctgggctcctgagcaccagaaggcatttgatg 3552
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3479 cgctaaccaaagaaaaaggggattctcctgggctcctgagcaccagaaggcatttgatg 3538

Qy  3553 ctatcaaaaaggccctgctgagcgcacctgctctggccctccctgacgtaactaaaccct 3612
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3539 ctatcaaaaaggccctgctgagcgcacctgctctggccctccctgacgtaactaaaccct 3598

Qy  3613 ttacccttttatgtggatgagcgtaagggagtagcccgaggagttttaacccaaaccctag 3672
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3599 ttacccttttatgtggatgagcgtaagggagtagcccgaggagttttaacccaaaccctag 3658

Qy  3673 gaccatggaggagacctgttgcctacctgtcaaagaagcttgatcctgtagccagtggtt 3732
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3659 gaccatggaggagacctgttgcctacctgtcaaagaagcttgatcctgtagccagtggtt 3718

Qy  3733 ggcccgtatgtctgaaggctatcgcagctgtggccatactggtcaaggacgctgacaaat 3792
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3719 ggcccgtatgtctgaaggctatcgcagctgtggccatactggtcaaggacgctgacaaat 3778

Qy  3793 tgactttgggacagaatataactgtaatagcccccatgcattggagaacatcgttcggc 3852
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3779 tgactttgggacagaatataactgtaatagcccccatgcattggagaacatcgttcggc 3838

Qy  3853 agccccagaccgatggatgaccaacgcccgcatgacccactatcaaagcctgcttctca 3912
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3839 agccccagaccgatggatgaccaacgcccgcatgacccactatcaaagcctgcttctca 3898

Qy  3913 cagagagggtcactttcgctccaccagccgctctcaaccctgccactcttctgcctgaag 3972
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3899 cagagagggtcactttcgctccaccagccgctctcaaccctgccactcttctgcctgaag 3958

Qy  3973 agactgatgaaccagtgactcatgattgccatcaactattgattgaggagactggggtcc 4032
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3959 agactgatgaaccagtgactcatgattgccatcaactattgattgaggagactggggtcc 4018

Qy  4033 gcaaggaccttacagacataccgctgactggagaagtgctaacctggttcactgacggaa 4092
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4019 gcaaggaccttacagacataccgctgactggagaagtgctaacctggttcactgacggaa 4078

Qy  4093 gcagctatgtggtggaaggtaagaggatggctggggcggcagtggtggacgggacccgca 4152
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4079 gcagctatgtggtggaaggtaagaggatggctggggcggcagtggtggacgggacccgca 4138

Qy  4153 cgatctgggccagcagcctgccggaaggaacttcagcgcaaaaggctgagctcatggccc 4212
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4139 cgatctgggccagcagcctgccggaaggaacttcagcgcaaaaggctgagctcatggccc 4198

Qy  4213 tcacgcaagctttgcggctggccgaagggaaatccataaacatttatacggacagcaggt 4272
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4199 tcacgcaagctttgcggctggccgaagggaaatccataaacatttatacggacagcaggt 4258

Qy  4273 atgcctttgcgactgcacacgtacacggggccatctataaacaaaggggttgcttacct 4332
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4259 atgcctttgcgactgcacacgtacacggggccatctataaacaaaggggttgcttacct 4318

Qy  4333 cagcagggagggaaataaagaacaaagaggaaattctaagcctattagaagccttacatt 4392
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4319 cagcagggagggaaataaagaacaaagaggaaattctaagcctattagaagccttacatt 4378

Qy  4393 tgccaaaaaggctagctattatacactgtcctggacatcagaaagccaaagatctcatat 4452
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4379 tgccaaaaaggctagctattatacactgtcctggacatcagaaagccaaagatctcatat 4438

Qy  4453 ctagagggaaccagatggctgaccgggttgccaagcaggcagcccaggctgttaaccttc 4512
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4439 ctagagggaaccagatggctgaccgggttgccaagcaggcagcccaggctgttaaccttc 4498

Qy  4513 tgcctataatagaaacgcccaaagccccagaacccagacgacagtacaccctagaagact 4572
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4499 tgcctataatagaaacgcccaaagccccagaacccagacgacagtacaccctagaagact 4558
```

```
Qy   4573 ggcaagagataaaaaagatagaccagttctctgagactccggaggggacctgctatacct 4632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4559 ggcaagagataaaaaagatagaccagttctctgagactccggaggggacctgctatacct 4618

Qy   4633 catatgggaaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatc 4692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4619 catatgggaaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatc 4678

Qy   4693 gtctaacccacctaggaactaaacacctgcagcagttggtcagaacatccccttatcatg 4752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4679 gtctaacccacctaggaactaaacacctgcagcagttggtcagaacatccccttatcatg 4738

Qy   4753 ttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctgg 4812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4739 ttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctgg 4798

Qy   4813 ttaatgctaatccttccagaatacctccaggaaagagactaaggggaagccacccaggcg 4872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4799 ttaatgctaatccttccagaatacctccaggaaagagactaaggggaagccacccaggcg 4858

Qy   4873 ctcactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatctat 4932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4859 ctcactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatctat 4918

Qy   4933 tggtttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagactt 4992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4919 tggtttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagactt 4978

Qy   4993 caaccgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaagg 5052
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4979 caaccgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaagg 5038

Qy   5053 taatagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaaga 5112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5039 taatagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaaga 5098

Qy   5113 tattggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggtag 5172
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5099 tattggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggtag 5158

Qy   5173 agaggatgaatagaaccattaaagagacccttaccaaattgaccacagagactggcatta 5232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5159 agaggatgaatagaaccattaaagagacccttaccaaattgaccacagagactggcatta 5218

Qy   5233 atgattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacagt 5292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5219 atgattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacagt 5278

Qy   5293 ttgggctgacccccctatgaattgctctacgggggaccccccccgttggcagaaattgcct 5352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5279 ttgggctgacccccctatgaattgctctacgggggaccccccccgttggcagaaattgcct 5338

Qy   5353 ttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcg 5412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5339 ttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcg 5398

Qy   5413 agtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagacttgc 5472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5399 agtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagacttgc 5458

Qy   5473 aagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaa 5532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5459 aagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaa 5518

Qy   5533 acctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacggctgtga 5592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5519 acctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacggctgtga 5578

Qy   5593 aagtcgaaggaatccccacctggatccatgcatcccacgttaagccggcgccacctcccg 5652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5579 aagtcgaaggaatccccacctggatccatgcatcccacgttaagcyggcgccacctcccg 5638

Qy   5653 actcggggtggagagccgaaaagactgagaatccccttaagcttcgcctccatcgcctgg 5712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5639 actcggggtggagagccgaaaagactgagaatccccttaagcttcgcctccatcgcctgg 5698

Qy   5713 ttccttactctaacaataactccccaggccagtagtaaacgccttatagacagctcgaac 5772
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5699 ttccttactctaacaataactccccaggccagtagtaaacgccttatagacagctcgaac 5758

Qy   5773 ccccatagacctttatcccttacctggctgattattgaccctgatacgggtgtcactgta 5832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5759 ccccatagacctttatcccttacctggctgattattgaccctgatacgggtgtcactgta 5818

Qy   5833 aatagcactcgaggtgttgctcctagaggcacctggtggcctgaactgcatttctgcctc 5892
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5819 aatagcactcgaggtgttgctcctagaggcacctggtggcctgaactgcatttctgcctc 5878

Qy   5893 cgattgattaaccccgctgttaaaagcacacctcccaacctagtccgtagttatgggttc 5952
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   5879 cgattgattaaccccgctgttaaragcacacctcccaacctagtccgtagttatgggttc 5938

Qy   5953 tattgctgcccaggcacagagaaagagaaatactgtgggggttctggggaatccttctgt 6012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5939 tattgctgcccaggcacagagaaagagaaatactgtgggggttctggggaatccttctgt 5998

Qy   6013 aggagatggagctgcgtcacctccaacgatggagactggaaatggccgatctctctccag 6072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5999 aggagatggagctgcgtcacctccaacgatggagactggaaatggccgatctctctccag 6058

Qy   6073 gaccgggtaaaattctcctttgtcaattccggcccgggcaagtacaaaatgatgaaacta 6132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6059 gaccgggtaaaattctcctttgtcaattccggcccgggcaagtacaaaatgatgaaacta 6118

Qy   6133 tataaagataagagctgctccccatcagacttagattatctaaagataagtttcactgaa 6192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6119 tataaagataagagctgctccccatcagacttagattatctaaagataagtttcactgaa 6178

Qy   6193 aggaaaacaggaaaatattcaaaagtggataaatggtatgagctggggaatagttttttta 6252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6179 aggaaaacaggaaaatattcaaaagtggataaatggtatgagctggggaatagttttttta 6238

Qy   6253 ttatatggcgggggagcagggtccactttaaccattcgccttaggatagagacggggaca 6312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6239 ttatatggcgggggagcagggtccactttaaccattcgccttaggatagagacggggaca 6298

Qy   6313 gaaccccctgtggcaatgggacccgataaagtactggctgaacaggggccccccggccctg 6372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6299 gaaccccctgtggcaatgggacccgataaagtactggctgaacaggggccccccggccctg 6358

Qy   6373 gagccaccgcataacttgccggtgccccaattaacctcgctgcggcctgacataacacag 6432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6359 gagccaccgcataacttgccggtgccccaattaacctcgctgcggcctgacataacacag 6418

Qy   6433 ccgcctagcaacagtaccactggattgattcctaccaacacgcctagaaactccccaggt 6492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6419 ccgcctagcaacagtaccactggattgattcctaccaacacgcctagaaactccccaggt 6478

Qy   6493 gttcctgttaagacaggacagagactcttcagtctcatccagggagctttccaagccatc 6552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6479 gttcctgttaagacaggacagagactcttcagtctcatccagggagctttccaagccatc 6538

Qy   6553 aactccaccgaccctgatgccacttcttcttgttggctttgtctatcctcagggcctcct 6612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6539 aactccaccgaccctgatgccacttcttcttgttggctttgtctatcctcagggcctcct 6598

Qy   6613 tattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaa 6672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6599 tattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaa 6658

Qy   6673 tgtacatgggggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgc 6732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6659 tgtacatgggggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgc 6718

Qy   6733 ataggaaaagctcccccatcccaccaacacctttgctatagtactgtggtttatgagcag 6792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    6719 ataggaaaagctccccccatcccaccaacacctttgctatagtactgtggtttatgagcag 6778

Qy    6793 gcctcagaaaatcagtatttagtacctggttataacaggtggtgggcatgcaatactggg 6852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6779 gcctcagaaaatcagtatttagtacctggttataacaggtggtgggcatgcaatactggg 6838

Qy    6853 ttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttctgtgtcatggtc 6912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6839 ttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttctgtgtcatggtc 6898

Qy    6913 caaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactat 6972
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6899 caaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactat 6958

Qy    6973 cggtataaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggatta 7032
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6959 cggtataaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggatta 7018

Qy    7033 gggacggccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagcta 7092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7019 gggacggccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagcta 7078

Qy    7093 gagaaaggacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggag 7152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7079 gagaaaggacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggag 7138

Qy    7153 tctgttagcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccgg 7212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7139 tctgttagcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccgg 7198

Qy    7213 aggggattagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaa 7272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7199 aggggattagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaa 7258

Qy    7273 tgttgcttctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaa 7332
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7259 tgttgcttctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaa 7318

Qy    7333 aagttagaggaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttc 7392
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7319 aagttagaggaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttc 7378

Qy    7393 aacaggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctg 7452
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7379 aacaggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctg 7438

Qy    7453 ctcctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaa 7512
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7439 ctcctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaa 7498

Qy    7513 cgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaa 7572
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7499 cgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaa 7558

Qy    7573 ggagaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaag 7632
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7559 ggagaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaag 7618

Qy    7633 tggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaa 7692
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7619 tggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaa 7678

Qy    7693 agctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtagaaggtca 7752
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    7679 agctctaaatgcccccgaattmcagaccctgctggctgccagtaaataggtagaaggtca 7738

Qy    7753 cacttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagtt 7812
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7739 cacttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagtt 7798

Qy    7813 gactaatcgcttatctggattctgtaaaactgactggcaccatagaagaattgattacac 7872
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7799,gactaatcgcttatctggattctgtaaaactgactggcaccatagaagaattgattacac 7858

Qy    7873 attgacagccctagtgacctatctcaactgcaatctgtcactctgcccaggagcccacgc 7932
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7859 attgacagccctagtgacctatctcaactgcaatctgtcactctgcccaggagcccacgc 7918

Qy    7933 agatgcggacctccggagctattttaaaatgattggtccacggagcgcgggctctcgata 7992
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7919 agatgcggacctccggagctattttaaaatgattggtccacggagcgcgggctctcgata 7978

Qy    7993 ttttaaaatgattggtccatggagcgcgggctctcgatattttaaaatgattggtttgtg 8052
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7979 ttttaaaatgattggtccatggagcgcgggctctcgatattttaaaatgattggtttgtg 8038

Qy    8053 acgcacaggctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgca 8112
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8039 acgcacaggctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgca 8098

Qy    8113 gtcctctacccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcct 8172
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8099 gtcctctacccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcct 8158

Qy    8173 cttgctgtttgcatcaaaaaaaaaaaaaaaaaaaaaaa 8209
           |||||||||||||||||||||||||||||||||||||
Db    8159 cttgctgtttgcatcaaaaaaaaaaaaaaaaaaaaaaa 8195
```

RESULT 3
AAF77726
ID AAF77726 standard; DNA; 7333 BP.
XX
AC AAF77726;
XX
DT 23-MAY-2001 (first entry)
XX
DE Defective retroviral genome isolated from PK-15 cell line.
XX
KW Retrovirus; graft transplantation; xenotransplantation; PK-15 cell line;
KW ss.
XX
OS Unidentified.
XX
PN US6190861-B1.
XX
PD 20-FEB-2001.
XX
PF 13-DEC-1996; 96US-0766528.
XX
PR 14-DEC-1995; 95US-0572645.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Fishman JA;
XX
DR WPI; 2001-256211/26.
DR P-PSDB; AAB73282, AAB73283, AAB73284.
XX
PT Assessing risk of endogenous retroviruses in clinical practice and in
PT xenotransplantation, comprises using probe sequences derived from swine
PT or miniature swine retroviral genome -
XX
PS Claim 1; Fig 2; 127pp; English.
XX
CC The present invention relates to a method for screening a cell or tissue
CC for the presence or expression of a retrovirus (RV), comprising
CC contacting a target nucleic acid from the cell or tissue with a second
CC nucleic acid from the present invention (e.g. the present sequence or a
CC fragment thereof). The method is useful for RV detection and to assess
CC graft transplantation risk. Screening of animals allows the elimination
CC of donors with active replication of known viruses. Inactive proviruses
CC can be detected and inactivated, allowing identification and elimination
CC of potential human pathogens derived from swine in a manner not possible
CC in the outbred human organ donor population and is important to the
CC development of human xenotransplantation.

XX
SQ Sequence 7333 BP; 1984 A; 1788 C; 1885 G; 1676 T; 0 other;

Query Match 36.1%; Score 2967; DB 22; Length 7333;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3507; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

```
Qy       1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
Db      11 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 70

Qy      61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggat 120
Db      71 aagaccgcttcttgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggat 130

Qy     121 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccaccccctta 180
Db     131 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccaccccctta 190

Qy     181 cacccgagaaccgacttggaggtaaagggatccccttggaacgtgtgtgtgtcggcc 240
Db     191 cacccgagaaccgacttggaggtaaagggatccccttggaacatatgtgtgtcggcc 250

Qy     241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
Db     251 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 310

Qy     301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
Db     311 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 370

Qy     361 ccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 420
Db     371 ccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 430

Qy     421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 480
Db     431 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 490

Qy     481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
Db     491 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 550

Qy     541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
Db     551 ctgtttcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 610

Qy     601 tgactacccccctagtttgactctcgaccattggactgaagttagatccagggctcata 660
Db     611 tgactacccccctagtttgactctcgaccattggactgaagttagatccagggctcata 670

Qy     661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
Db     671 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 730

Qy     721 cattcgatgttggatggccatcagaggggaccttaattctgaaattatcctggctgtta 780
Db     731 cattcgatgttggatggccatcagaggggaccttaattctgaaattatcctggctgtta 790

Qy     781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 840
Db     791 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 850

Qy     841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
Db     851 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 910

Qy     901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
Db     911 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 970

Qy     961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
Db     971 tcgagccctc-tcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1029

Qy    1021 caacctgttccccccacccccttatccagcacagggtgctgtgaggggga-cctctgcccct 1079
Db    1030 caacctgttccccccacccccttatccagcacagggtgctgtgaggggaccctctgcccct 1089

Qy    1080 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1139
Db    1090 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1149

Qy    1140 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1199
Db    1150 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1209

Qy    1200 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1259
Db    1210 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1269

Qy    1260 aaaactaaccatccccctttctcggaggatccccaacgcctcacgggttggtggagtcc 1319
Db    1270 aaaactaaccatccccctttctcggaggatccccaacgcctcacgggttggtggagtcc 1329

Qy    1320 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1379
Db    1330 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1389

Qy    1380 acaaccgaggagcgagagagaattctgttagaggctagaaaaaaatgttcctggggccgac 1439
Db    1390 acaaccgaggagcgagagagaattctgttagaggctaaaaaaaatgttcctggggccgac 1449

Qy    1440 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1499
Db    1450 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1509

Qy    1500 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1559
Db    1510 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1569

Qy    1560 gcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatg 1619
Db    1570 gcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatg 1629

Qy    1620 cagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcgg 1679
Db    1630 cagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcgg 1689

Qy    1680 ttcaccccttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcatt 1739
Db    1690 ttcaccccttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcatt 1749

Qy    1740 gggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggct 1799
Db    1750 gggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggct 1809

Qy    1800 gagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagag 1859
Db    1810 gagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagag 1869

Qy    1860 gagaaggaacagagaaaagaaaaggagagagaagaaaggggaggaaagacgtgatagacgg 1919
Db    1870 gagaaggaacagagaaaagaaaaggagagagaagaaaggggaggaaagacgtgatagacgg 1929

Qy    1920 caagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggag 1979
Db    1930 caagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggag 1989

Qy    1980 agagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaat 2039
Db    1990 agagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaat 2049

Qy    2040 aggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagg 2099
```

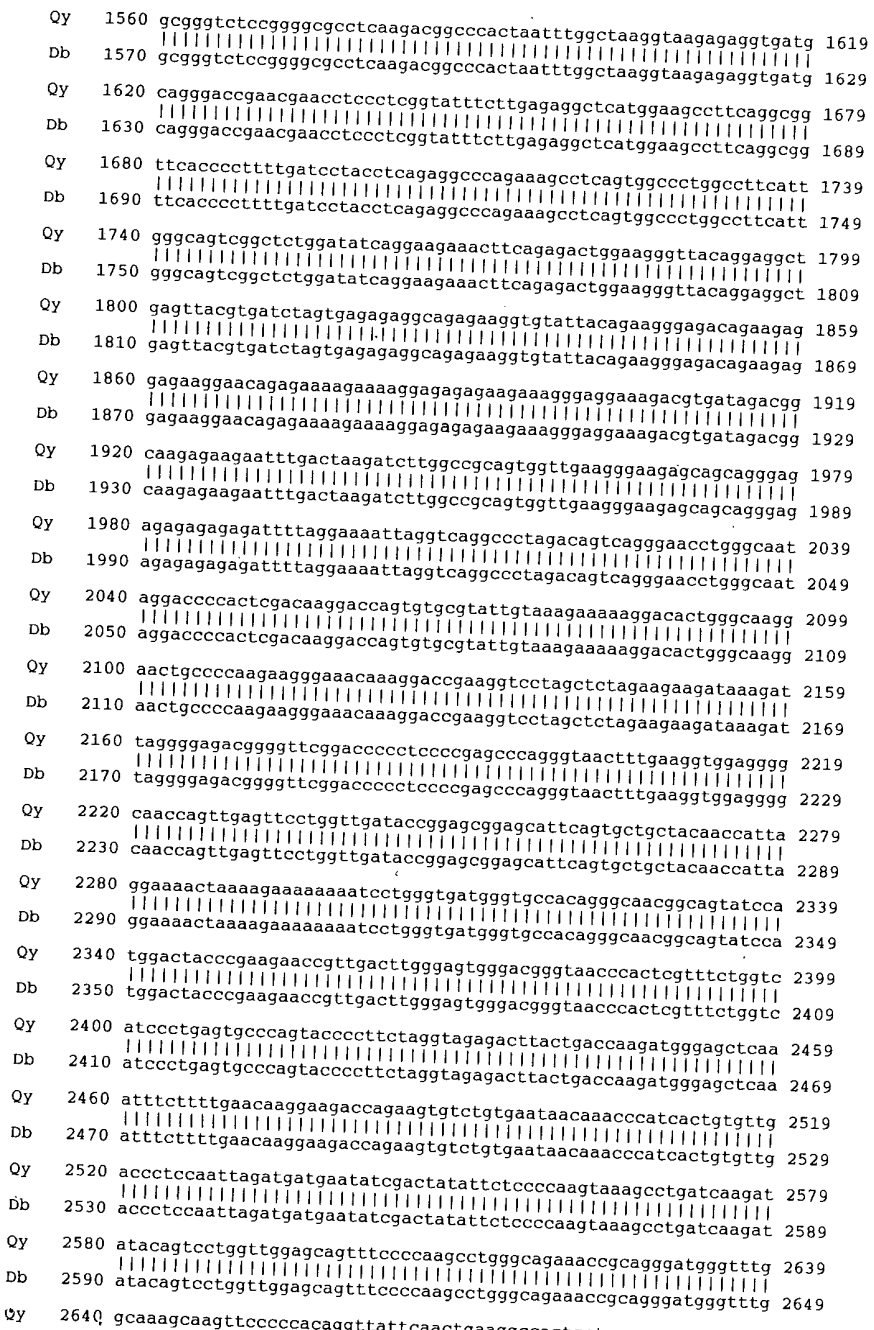
```
Db    2050 aggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagg 2109

Qy    2100 aactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagat 2159
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2110 aactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagat 2169

Qy    2160 taggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtggagggg 2219
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2170 taggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtggagggg 2229

Qy    2220 caaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccatta 2279
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2230 caaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccatta 2289

Qy    2280 ggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatcca 2339
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2290 ggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatcca 2349

Qy    2340 tggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtc 2399
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2350 tggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtc 2409

Qy    2400 atccctgagtgcccagtacccttctaggtagagacttactgaccaagatgggagctcaa 2459
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2410 atccctgagtgcccagtacccttctaggtagagacttactgaccaagatgggagctcaa 2469

Qy    2460 atttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttg 2519
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2470 atttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttg 2529

Qy    2520 accctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagat 2579
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2530 accctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagat 2589

Qy    2580 atacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttg 2639
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2590 atacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttg 2649

Qy    2640 gcaaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtc 2699
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2650 gcaaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtc 2709

Qy    2700 agacagtacccccttgagtagagaggctcgagaaggaatttggccgcatgttcaaagatta 2759
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2710 agacagtacccccttgagtagagaggctcgagaaggaatttggccgcatgttcaaagatta 2769

Qy    2760 atccaacagggcatcctagttcctgtccaatccccttggaatactccccctgctaccggtt 2819
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2770 atccaacagggcatcctagttcctgtccaatccccttggaatactccccctgctaccggtt 2829

Qy    2820 aggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaagg 2879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2830 aggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaagg 2889

Qy    2880 gtgcaggacatacacccaacggtcccgaaccttataacctcttgagcgccctcccgcct 2939
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2890 gtgcaggacatacacccaacggtcccgaaccttataacctcttgagcgccctcccgcct 2949

Qy    2940 gaacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacac 2999
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2950 gaacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacac 3009

Qy    3000 cccactagccaaccacttttttgccttcgaatggagagatccaggtacgggaagaaccggg 3059
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    3010 cccactagccaaccacttttttaccttcgaatggagagatccaggtacgggaagaaccggg 3069

Qy    3060 cagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaa 3119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3070 cagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaa 3129

Qy    3120 gccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccag 3179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3130 gccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccag 3189

Qy    3180 tacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaag 3239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3190 tacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaag 3249

Qy    3240 gcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagatt 3299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3250 gcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagatt 3309

Qy    3300 tgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacg 3359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3310 tgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacg 3369

Qy    3360 gaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgaga 3419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3370 gaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgaga 3429

Qy    3420 gagtttttggggacagctggattttgcagactgtggatcccggggtttgcgaccttagca 3479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3430 gagtttttggggacagctggattttgcagactgtggatcccggggtttgcgaccttagca 3489

Qy    3480 gccccactctacccgctaaccaaagaaaaaggggg 3514
           ||||||||||||||||||||||||||||||||||
Db    3490 gccccactctacccgctaaccaaagaaaaaggggg 3524
```

```
RESULT   4
AAT74883
ID   AAT74883 standard; cDNA; 7393 BP.
XX
AC   AAT74883;
XX
DT   09-FEB-1998  (first entry)
XX
DE   Porcine retrovirus cDNA (defective).
XX
KW   Retrovirus; porcine; GAG protein; POL protein; ENV protein;
KW   xenotransplantation; infectious; provirus; organ transplant; donor;
KW   activated virus; PCR; ss.
XX
OS   Porcine retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             598..2172
FT                   /*tag=  a
FT   mat_peptide     598..2169
FT                   /*tag=  b
FT                   /note= "putative GAG protein"
FT   CDS             2320..4737
FT                   /*tag=  c
FT                   /note= "putative POL coding region (partial) as
FT                        described in the specification"
FT   mat_peptide     2320..3522
FT                   /*tag=  d
FT                   /note= "putative POL protein (partial)"
FT   mat_peptide     3516..4328
FT                   /*tag=  e
FT                   /note= "putative POL protein (partial)"
FT   mat_peptide     4332..4748
FT                   /*tag=  f
FT   CDS             4738..6725
FT                   /*tag=  g
FT                   /note= "putative ENV coding region (partial) as
FT                        described in the specification"
FT   mat_peptide     4752..6722
FT                   /*tag=  h
FT                   /note= "ENV protein (partial)"
XX
PN   WO9721836-A1.
XX
PD   19-JUN-1997.
XX
PF   13-DEC-1996;   96WO-US19680.
```

XX
PR 14-DEC-1995; 95US-0572645.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Fishman JA;
XX
DR WPI; 1997-332804/30.
DR P-PSDB; AAW32091-W32095.
XX
PT New nucleic acid from porcine retro:viruses - used for detecting
PT viruses in transplant or other tissue and for assessing risk of
PT transmitting infection to graft recipient
XX
PS Claim 16; Fig 2; 128pp; English.
XX
CC This cDNA sequence represents a defective purified swine retrovirus
CC found in PK-15 cells containing the putative coding regions for viral
CC GAG, POL and ENV proteins. There are a few in frame stop codons and
CC apparent frame shifts in the given coding sequence which alter features
CC of the translation. This sequence and PCR fragments generated from the
CC sequence (see AAT74812-T74882) could be used to screen organs for the
CC presence of porcine retroviruses prior to xenotransplantation.
CC Transplantation can increase the likelihood of retroviral activation if
CC intact and infectious proviruses are present. The porcine retroviral
CC sequence can be used to generate probes to determine the level (e.g.
CC copy number) of intact (i.e. potentially replicating) porcine provirus
CC sequences in a strain of xenograft transplantation donors. It can be
CC used to detect mutations, genetic lesions or viral recombinants and to
CC determine the histological localisation of activated retrovirus. Using
CC Polymerase Chain Reaction DNA Quantitation (PDQ) on blood mononuclear
CC cells, infectivity titration and susceptibility testing can be
CC performed. Ultimately animal donors without intact porcine retroviral
CC sequences or a lower copy number of viral elements could be selected.
XX
SQ Sequence 7393 BP; 2004 A; 1801 C; 1896 G; 1692 T; 0 other;

Query Match 36.1%; Score 2967; DB 18; Length 7393;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3507; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

```
Qy       1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
Db      11 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 70

Qy      61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggat 120
Db      71 aagaccgcttcttgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggat 130

Qy     121 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctta 180
Db     131 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctta 190

Qy     181 cacccgagaaccgacttggaggtaaagggatccccttggaacgtgtgtgtgtcggcc 240
Db     191 cacccgagaaccgacttggaggtaaagggatccccttggaacatatgtgtgtgtcggcc 250

Qy     241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
Db     251 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 310

Qy     301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
Db     311 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 370

Qy     361 ccacccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 420
Db     371 ccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 430

Qy     421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 480
Db     431 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 490

Qy     481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
Db     491 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 550

Qy     541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
Db     551 ctgtttcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 610

Qy     601 tgactacccccctagtttgactctcgaccattggactgaagttagatccagggctcata 660
Db     611 tgactacccccttagtttgactctcgaccattggactgaagttagatccagggctcata 670

Qy     661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
Db     671 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 730

Qy     721 cattcgatgttggatggccatcagaggggacctttaattctgaaattatcctggctgtta 780
Db     731 cattcgatgttggatggccatcagaggggacctttaattctgaaattatcctggctgtta 790

Qy     781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 840
Db     791 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcctta 850

Qy     841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
Db     851 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 910

Qy     901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
Db     911 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 970

Qy     961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
Db     971 tcgagccctc-tcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1029

Qy    1021 caacctgttcccccaccccctttatccagcacagggtgctgtgaggggacctctgcccct 1089
Db    1030 caacctgttcccccaccccctttatccagcacagggtgctgtgaggggacctctgcccct 1089

Qy    1080 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1139
Db    1090 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1149

Qy    1140 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1199
Db    1150 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1209

Qy    1200 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1259
Db    1210 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1269

Qy    1260 aaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtcc 1319
Db    1270 aaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtcc 1329

Qy    1320 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1379
Db    1330 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1389

Qy    1380 acaaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgac 1439
Db    1390 acaaccgaggagcgagagagaattctgttagaggctaaaaaaaatgttcctggggccgac 1449

Qy    1440 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1499
Db    1450 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1509

Qy    1500 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1559
Db    1510 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1569
```



RESULT 5
AAV09698
ID AAV09698 standard; DNA; 3320 BP.
XX
AC AAV09698;
XX
DT 19-MAY-1998 (first entry)
XX
DE Porcine retrovirus pol and env DNA.
XX
KW Porcine retrovirus; PoEV; POL protein; ENV protein; vaccine;
KW diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
XX

```
OS   Porcine retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             23..2793
FT                   /*tag=  a
FT                   /product= POL protein
FT                   /note= "polymerase protein"
FT   CDS             2642..3297
FT                   /*tag=  b
FT                   /product= ENV protein
FT                   /note= "envelope protein"
XX
PN   WO9740167-A1.
XX
PD   30-OCT-1997.
XX
PF   18-APR-1997;   97WO-GB01087.
XX
PR   10-FEB-1997;   97GB-0002668.
PR   19-APR-1996;   96GB-0008164.
XX
PA   (IMUT-) IMUTRAN LTD.
PA   (QONE-) Q-ONE BIOTECH LTD.
XX
PI   Galbraith DN,  Haworth C,  Lees GM,  Smith KT;
XX
DR   WPI; 1997-535851/49.
XX
PT   Polynucleotide encoding porcine retrovirus expression product -
PT   useful to develop products for use in vaccines, diagnosis and
PT   xeno-transplantation
XX
PS   Claim 4; Fig 1; 69pp; English.
XX
CC   This DNA sequence encodes the porcine retrovirus (PoEV) polymerase (POL)
CC   and envelope (ENV) proteins. These proteins can be used to develop viral
CC   vaccines, antisense nucleic acids, ribozymes and other antiviral agents.
CC   They can also be used in xeno-transplantation technology and as
CC   diagnostic tools.
XX
SQ   Sequence 3320 BP; 922 A; 844 C; 832 G; 722 T; 0 other;
```

Query Match 32.7%; Score 2682; DB 18; Length 3320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   2972 agatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcgaatg 3031
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 agatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcgaatg 80

Qy   3032 gagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggtt 3091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 gagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggtt 140

Qy   3092 caagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggat 3151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 caagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggat 200

Qy   3152 ccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccac 3211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 ccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccac 260

Qy   3212 caaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggcta 3271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 caaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggcta 320

Qy   3272 cagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacag 3331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 cagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacag 380

Qy   3332 tttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagatacc 3391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 tttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagatacc 440

Qy   3392 ggccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagact 3451
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 ggccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagact 500

Qy   3452 gtggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaagg 3511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 gtggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaagg 560

Qy   3512 gggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgct 3571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 gggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgct 620

Qy   3572 gagcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatga 3631
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621 gagcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatga 680

Qy   3632 gcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgt 3691
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    681 gcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgt 740

Qy   3692 tgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggc 3751
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    741 tgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggc 800

Qy   3752 tatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatat 3811
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    801 tatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatat 860

Qy   3812 aactgtaatagcccccccatgcattggagaacatcgttcggcagccccccagaccgatggat 3871
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    861 aactgtaatagcccccccatgcattggagaacatcgttcggcagccccccagaccgatggat 920

Qy   3872 gaccaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgc 3931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    921 gaccaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgc 980

Qy   3932 tccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgac 3991
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    981 tccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgac 1040

Qy   3992 tcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacat 4051
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1041 tcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacat 1100

Qy   4052 accgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaagg 4111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1101 accgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaagg 1160

Qy   4112 taagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcct 4171
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1161 taagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcct 1220

Qy   4172 gccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggct 4231
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1221 gccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggct 1280

Qy   4232 ggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcaca 4291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1281 ggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcaca 1340

Qy   4292 cgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaa 4351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1341 cgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaa 1400

Qy   4352 gaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctat 4411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1401 gaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctat 1460

Qy   4412 tatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggc 4471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1461 tatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggc 1520
```

```
Qy   4472 tgaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcc 4531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1521 tgaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcc 1580

Qy   4532 caaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagat 4591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1581 caaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagat 1640

Qy   4592 agaccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaatcct 4651
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1641 agaccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaatcct 1700

Qy   4652 gccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaac 4711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1701 gccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaac 1760

Qy   4712 taaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagt 4771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1761 taaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagt 1820

Qy   4772 ggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccag 4831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1821 ggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccag 1880

Qy   4832 aatacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggactt 4891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1881 aatacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggactt 1940

Qy   4892 cactgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctt 4951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1941 cactgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctt 2000

Qy   4952 ttcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaa 5011
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2001 ttcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaa 2060

Qy   5012 aatactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatgg 5071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2061 aatactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatgg 2120

Qy   5072 tccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaa 5131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2121 tccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaa 2180

Qy   5132 actgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccat 5191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2181 actgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccat 2240

Qy   5192 taaagagacccttaccaaattgaccacagagactggcattaatgattggatggctctcct 5251
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2241 taaagagacccttaccaaattgaccacagagactggcattaatgattggatggctctcct 2300

Qy   5252 gccctttgtgcttttagggtgaggaacacccctggacagtttgggctgacccccctatga 5311
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2301 gccctttgtgcttttagggtgaggaacacccctggacagtttgggctgacccccctatga 2360

Qy   5312 attgctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctgatgt 5371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2361 attgctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctgatgt 2420

Qy   5372 gctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcgagc 5431
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2421 gctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcgagc 2480

Qy   5432 gtggaagcagctccgggaggcctactcaggaggagacttgcaagttccacatcgcttcca 5491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2481 gtggaagcagctccgggaggcctactcaggaggagacttgcaagttccacatcgcttcca 2540

Qy   5492 agttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaa 5551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2541 agttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaa 2600
```

```
Qy   5552 gggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatccccac 5611
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2601 gggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatccccac 2660

Qy   5612 ctggatccatgcatcccacgttaagccggcgccacctcccga 5653
          |||||||||||||||||||||||||||||||||||||||||
Db   2661 ctggatccatgcatcccacgttaagccggcgccacctcccga 2702
```

RESULT 6
AAV09703
ID AAV09703 standard; DNA; 1974 BP.
XX
AC AAV09703;
XX
DT 20-MAY-1998 (first entry)
XX
DE Porcine retrovirus Raji clone ENV DNA.
XX
KW Porcine retrovirus; PoEV; ENV protein; envelope protein; vaccine;
KW diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
XX
OS Porcine retrovirus.
XX
FH Key Location/Qualifiers
FT CDS 1..1974
FT /*tag= a
FT /product= ENV protein
FT /note= "envelope protein"
XX
PN WO9740167-A1.
XX
PD 30-OCT-1997.
XX
PF 18-APR-1997; 97WO-GB01087.
XX
PR 10-FEB-1997; 97GB-0002668.
PR 19-APR-1996; 96GB-0008164.
XX
PA (IMUT-) IMUTRAN LTD.
PA (QONE-) Q-ONE BIOTECH LTD.
XX
PI Galbraith DN, Haworth C, Lees GM, Smith KT;
XX
DR WPI; 1997-535851/49.
DR P-PSDB; AAW39274.
XX
PT Polynucleotide encoding porcine retrovirus expression product -
PT useful to develop products for use in vaccines, diagnosis and
PT xeno-transplantation
XX
PS Claim 6; Fig 4; 69pp; English.
XX
CC This sequence encodes the porcine retrovirus envelope (ENV) protein
CC isolated from the human cell line Raji. Such viral proteins can be used
CC to develop viral vaccines, antisense nucleic acids, ribozymes and other
CC antiviral agents. They can also be used in xeno-transplantation
CC technology and as diagnostic tools.
XX
SQ Sequence 1974 BP; 534 A; 489 C; 483 G; 468 T; 0 other;

Query Match 18.3%; Score 1499; DB 18; Length 1974;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

```
Qy   5639 ggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcg 5698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 ggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcg 79

Qy   5699 cctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttа 5758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 cctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttа 139
```

```
Qy   5759 tagacagctcgaacccccatagacctttatcccttacctggctgattattgaccctgata 5818
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    140 tagacagctcgaacccccatagacctttatcccctacctggctgattattgaccctgata 199

Qy   5819 cgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac 5878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 cgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac 259

Qy   5879 tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtcc 5938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    260 tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtcc 319

Qy   5939 gtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctg 5998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 gtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctg 379

Qy   5999 gggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggc 6058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 gggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggc 439

Qy   6059 cgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtaca 6118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    440 cgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtaca 499

Qy   6119 aaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaaga 6178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    500 aaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaaga 559

Qy   6179 taagtttcactg--aaaggaaaacaggaaaatattcaaaagtggataaatggtatgagct 6236
          |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db    560 taagtttcactgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagct 619

Qy   6237 ggggaatagttttttat-tatatggcgggggagcagggtccactttaaccattcgccttа 6295
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    620 ggggaatagttttttataaatatggcgggggagcagggtccactttaaccattcgccttа 679

Qy   6296 ggatagagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaac 6355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680 ggatagagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaac 739

Qy   6356 aggggccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgc 6415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740 aggggccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgc 799

Qy   6416 ggcctgacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgc 6475
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    800 ggcctgacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgc 859

Qy   6476 ctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagg 6535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    860 ctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagg 919

Qy   6536 gagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtc 6595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    920 gagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtc 979

Qy   6596 tatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgacca 6655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    980 tatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgacca 1039

Qy   6656 aagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagttt 6715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1040 aagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagttt 1099

Qy   6716 ccgggaaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagta 6775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1100 ccgggaaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagta 1159

Qy   6776 ctgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggt 6835
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1160 ctgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggt 1219

Qy   6836 gggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaag 6895
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1220 gggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaag 1279

Qy   6896 atttctgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtcc 6955
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1280 atttgtgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtcc 1339

Qy   6956 ttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctag 7015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1340 ttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctag 1399

Qy   7016 ctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatca 7075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1400 ctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatca 1459

Qy   7076 caggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatc 7135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1460 caggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatc 1519

Qy   7136 tccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaag 7195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1520 tccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaag 1579

Qy   7196 tggttctacagaaccggagggggattagatctgctgtttctaagagaaggtgggttatgtg 7255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1580 tggttctacagaaccggagggggattagatctgctgtttctaagagaaggtgggttatgtg 1639

Qy   7256 cagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactcca 7315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1640 cagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactcca 1699

Qy   7316 tgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggt 7375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1700 tgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggt 1759

Qy   7376 ggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacgg 7435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1760 ggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacgg 1819

Qy   7436 ggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttg 7495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1820 ggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttg 1879

Qy   7496 ttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtacc 7555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1880 ttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtacc 1939

Qy   7556 aaggccttctgagccaaggagaaactgacctctag 7590
          |||||||||||||||||||||||||||||||||
Db   1940 aaggccttctgagccaaggagaaactgacctctag 1974
```

```
RESULT   7
AAV82749
ID   AAV82749 standard; DNA; 3482 BP.
XX
AC   AAV82749;
XX
DT   25-FEB-1999  (first entry)
XX
DE   Pig endogenous retrovirus (PERV)-B envelope (env) gene region.
XX
KW   Pig endogenous retrovirus; PERV-A; envelope protein; PERV-B; subtype;
KW   probe; primer; detection; retrovirus; human tissue; xenotransplant;
KW   primary porcine tissue; human cell line; porcine cell line; ss.
XX
OS   Pig endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             911..2884
FT                   /*tag=  a
FT                   /product= envelope_protein
```

XX
PN WO9853104-A2.
XX
PD 26-NOV-1998.
XX
PF 18-MAY-1998; 98WO-GB01428.
XX
PR 16-MAY-1997; 97GB-0010154.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Stoye JP, Weiss RA;
XX
DR WPI; 1999-045324/04.
DR P-PSDB; AAW85452.
XX
PT Newly isolated nucleic acid probe capable of hybridising to either
PT the PERV-A or PERV-B env gene - useful in the detection of
PT retroviruses, and their subtypes, in a sample of porcine/human
PT tissue
XX
PS Claim 3; Page 21-23; 36pp; English.
XX
CC The present sequence encodes a Pig endogenous retrovirus (PERV)-B
CC envelope protein. PERV exists in two different subtypes, PERV-A and
CC PERV-B. The differences are reflected in sequence divergence in the
CC envelope genes. Probes and primers can be derived from the envelope
CC (env) genes of PERV-A and PERV-B. The probes and primers are used in
CC a method to detect retroviruses in a sample of porcine/human tissue,
CC particularly primary porcine tissue and human cell lines that have been
CC cultivated in the presence of a porcine cell line, or human tissue from
CC a patient with a xenotransplant. Subtype of PERV in a sample containing
CC one of the PERV env genes can also be determined.
XX
SQ Sequence 3482 BP; 927 A; 854 C; 867 G; 834 T; 0 other;

Query Match 17.9%; Score 1466; DB 20; Length 3482;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2536; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

```
Qy   5639 ggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcg 5698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    930 ggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcg 989

Qy   5699 cctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttа 5758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    990 cctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttа 1049

Qy   5759 tagacagctcgaacccccatagacctttatcccttacctggctgattattgaccctgata 5818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1050 tagacagctcgaacccccatagacctttatcccttacctggctgattattgaccctgata 1109

Qy   5819 cgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac 5878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1110 cgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac 1169

Qy   5879 tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtcc 5938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtcc 1229

Qy   5939 gtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctg 5998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 gtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctg 1289

Qy   5999 gggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggc 6058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 gggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggc 1349

Qy   6059 cgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtaca 6118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1350 cgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtaca 1409

Qy   6119 aaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaaga 6178
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1410 aagtgatgaaactatataaagataagagctgctccccatcagacttagattatctaaaga 1469

Qy   6179 taagtttcactg--aaaggaaaacaggaaaatattcaaaagtggataaatggtatgagct 6236
          ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db   1470 taagtttcactgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagct 1529

Qy   6237 ggggaatagttttttat-tatatggcgggggagcagggtccactttaaccattcgcctta 6295
          ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db   1530 ggggaatagttttttataaatatggcgggggagcagggtccactttaaccattcgcctta 1589

Qy   6296 ggatagagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaac 6355
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   1590 ggatagagacggggacagaaccccctgtggcagtgggacccgataaagtactggctgaac 1649

Qy   6356 aggggccccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgc 6415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1650 aggggccccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgc 1709

Qy   6416 ggcctgacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgc 6475
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1710 ggcctgacataacacagccgcctagcaacggtaccactggattgattcctaccaacacgc 1769

Qy   6476 ctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagg 6535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1770 ctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagg 1829

Qy   6536 gagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtc 6595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1830 gagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtc 1889

Qy   6596 tatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgacca 6655
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1890 tatcctcagggcctccttattatgaggggatggctaaagaaggaaaattcaatgtgacca 1949

Qy   6656 aagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagttt 6715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1950 aagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagttt 2009

Qy   6716 ccgggaaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagta 6775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2010 ccgggaaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagta 2069

Qy   6776 ctgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggt 6835
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2070 ctgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggt 2129

Qy   6836 gggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaag 6895
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2130 gggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaag 2189

Qy   6896 atttctgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtcc 6955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2190 atttctgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtcc 2249

Qy   6956 ttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctag 7015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2250 ttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctag 2309

Qy   7016 ctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatca 7075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2310 ctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatca 2369

Qy   7076 caggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatc 7135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2370 caggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatc 2429

Qy   7136 tccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaag 7195
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2430 tccgagccttagaggagtctgttagcaacctagaagagtccctgacttctttgtctgaag 2489

Qy   7196 tggttctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtg 7255
```

```
Db    2490 tggttctacagaaccggagggattagatctgctgtttctaagagaaggtgggttatgtg 2549

Qy    7256 cagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactcca 7315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2550 cagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactcca 2609

Qy    7316 tgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggt 7375
           ||| |||||||||| ||| ||||||||||||||||||||||||||||||||||||||||
Db    2610 tgagcaagcttagagaaaggttagagaggcgtcgaagggaaagagaggctgaccaggggt 2669

Qy    7376 ggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacgg 7435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2670 ggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacgg 2729

Qy    7436 ggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttg 7495
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2730 gaccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttg 2789

Qy    7496 ttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtacc 7555
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2790 ttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtacc 2849

Qy    7556 aaggccttctgagccaaggagaaactgacctctagccttcccagttctaagattagaact 7615
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2850 aaggccttctgagccaaggagaaactgacctctagccttcccagttctaagattagaact 2909

Qy    7616 attaacaagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaac 7675
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2910 attaacaagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaac 2969

Qy    7676 ccaggaagttaataaaaagctctaaatgccccgaattacagaccctgctggctgccagt 7735
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    2970 ccaggaagttaataaaaagctctaaatgccccgaattccagaccctgctggctgccagt 3029

Qy    7736 aaataggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaag 7795
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3030 aaataggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaag 3089

Qy    7796 ataacaggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccat 7855
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    3090 ataacaggaaatgagttgactaatcgcttatctggattctgtaaaaccgactggcaccat 3149

Qy    7856 agaagaattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactc 7915
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3150 agaagaattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactc 3209

Qy    7916 tgcccaggagcccacgcagatgcggacctccggagctattttaaaatgattggtccacgg 7975
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3210 tgcccaggagcccacgcagatgcggacctccggagctattttaaaatgattggtccacgg 3269

Qy    7976 agcgcgggctctcgatattttaaaatgattggtccatggagcgcgggctc-tcgatattt 8034
           |||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||| 
Db    3270 agcgcgggctctcgatattttaaaatgattggtccacggagcgcgggctcttcgatattt 3329

Qy    8035 taaaatgattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgat 8094
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3330 taaaatgattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgat 3389

Qy    8095 tccgcactcggggccgcagtcctctaccccctgcgtggtgtacgactgtgggccccagcgc 8154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3390 tccgcactcggggccgcagtcctctaccccctgcgtggtgtacgactgtgggccccagcgc 3449

Qy    8155 gcttggaataaaaatcctcttgctgtttgcatc 8187
           |||||||||||||||||||||||||||||||
Db    3450 gcttggaataaaaatcctcttgctgtttgcatc 3482


RESULT   8
AAC67020
ID   AAC67020 standard; DNA; 4402 BP.
XX
AC   AAC67020;
XX
DT   27-MAR-2001  (first entry)
XX
DE   PERV env protein coding sequence SEQ ID NO: 20.
XX
KW   Xenotransplantation; infectious agent; vaccine; ds.
XX
OS   Porcine endogenous retrovirus.
XX
PN   WO200071726-A1.
XX
PD   30-NOV-2000.
XX
PF   24-MAY-2000; 2000WO-US14296.
XX
PR   24-MAY-1999;   99US-0135631.
XX
PA   (MAYO-) MAYO MEDICAL VENTURES.
XX
PI   Federspiel MJ;
XX
DR   WPI; 2001-032041/04.
XX
PT   Inhibiting or preventing infectious agent transmission in mammalian
PT   transplant recipients, by introducing recombinant DNA comprising DNA
PT   encoding extracellular proteins of the agent into donor cells, such as
PT   swine cells -
XX
PS   Claim 16; Page 105-106; 144pp; English.
XX
CC   The present invention provides a method to prevent the transmission of
CC   infectious agents during xenotransplantation. This involves introducing
CC   to donor swine cells a recombinant DNA encoding a peptide fragment from
CC   the infectious agent, and then introducing these cells into the
CC   transplant recipient.
XX
SQ   Sequence 4402 BP; 1259 A; 1085 C; 1111 G; 947 T; 0 other;


  Query Match             10.3%;  Score 848;  DB 22;  Length 4402;
  Best Local Similarity   98.8%;  Pred. No. 0;
  Matches 2198;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy    2982 ttctgcctgagattacaccccactagccaaccacttttgccttcgaatggagagatcca 3041
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1043 ttctgcctgagattacaccccactagccaaccacttttgccttcgaatggagagatcca 1102

Qy    3042 ggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactcc 3101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1103 ggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactcc 1162

Qy    3102 ccgaccatctttgacgaagccctacacagggacctggccaacttcaggatccaacaccct 3161
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    1163 ccgaccatctttgacgaagccctacacagagacctggccaacttcaggatccaacaccct 1222

Qy    3162 caggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggac 3221
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1223 caggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggac 1282

Qy    3222 tgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctacagagcctct 3281
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    1283 tgcttagaaggcacgaaggcactactgctggaattgtctgacctaggctacagagcctct 1342

Qy    3282 gctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggc 3341
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    1343 gctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcgggac 1402

Qy    3342 gggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacc 3401
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1403 gggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacc 1462

Qy    3402 acagccaaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccg 3461
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1463 acagccaaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccg 1522

Qy   3462 gggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggggggattctcc 3521
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||| |||||||
Db   1523 gggtttgcgaccttagcagccccactctacccactaaccaaagaaaaaggggaattctcc 1582

Qy   3522 tgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacct 3581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1583 tgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacct 1642

Qy   3582 gctctggccctccctgacgtaactaaacccttttaccctttatgtggatgagcgtaaggga 3641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1643 gctctggccctccctgacgtaactaaacccttttaccctttatgtggatgagcgtaaggga 1702

Qy   3642 gtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctg 3701
          |||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1703 gtagcccggggagttttaacccaaactctaggaccatggaggagacctgttgcctacctg 1762

Qy   3702 tcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagct 3761
          |||||||||| |||||||||||||||||||||||||||| |||||||||||||||||||
Db   1763 tcaaagaagctcgatcctgtagccagtggttggcccgtatgcctgaaggctatcgcagct 1822

Qy   3762 gtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaata 3821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1823 gtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaata 1882

Qy   3822 gccccccatgcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgcc 3881
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1883 gccccccatgcgttggagaacatcgttcggcagcccccagaccgatggatgaccaacgcc 1942

Qy   3882 cgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgctccaccagcc 3941
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   1943 cgcatgacccactatcaaagcctgcttctcacagagagggtcacgttcgctccaccagcc 2002

Qy   3942 gctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgc 4001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2003 gctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgc 2062

Qy   4002 catcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgact 4061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2063 catcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgact 2122

Qy   4062 ggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatg 4121
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2123 ggagaagtgttaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatg 2182

Qy   4122 gctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctgccggaagga 4181
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2183 gctggggcggcggtggtggacgggacccgcacgatctgggccagcagcctgccggaagga 2242

Qy   4182 acttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaaggg 4241
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2243 acttcagcacaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaaggg 2302

Qy   4242 aaatccataaacatttatacggacagcaggtatgcctttgcgactgcacacgtacacggg 4301
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   2303 aaatccataaacatttatacagacagcaggtatgcctttgcgactgcacacgtacacggg 2362

Qy   4302 gccatctataaacaaaggggttgcttacctcagcagggagggaaataaagaacaaagag 4361
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2363 gccatctataagcaaaggggttgcttacctcagcagggagggaaataaagaacaaagag 2422

Qy   4362 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 4421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2423 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 2482

Qy   4422 cctggacatcagaaagccaaagatctcatatctagagggaaccagatggctgaccgggtt 4481
          |||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||||
Db   2483 cctggacatcagaaagccaaagatcccatatccagagggaaccagatggctgaccgggtt 2542

Qy   4482 gccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaaagcccca 4541
          |||||||||||||||||| |||||||||||||||| |||||||| ||||||||||||
Db   2543 gccaagcaggcagcccagggtgttaaccttctgcctatgatagaaacacccaaagcccca 2602

Qy   4542 gaacccagacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttc 4601
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2603 gaacccggacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttc 2662

Qy   4602 tctgagactccggaggggacctgctatacctcatatgggaaggaaatcctgccccacaaa 4661
          |||||||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||
Db   2663 tctgagactccggaagggacctgctatacctcagatgggaaggaaatcctgccccacaaa 2722

Qy   4662 gaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaaacacctg 4721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2723 gaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaaacacctg 2782

Qy   4722 cagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctgactcg 4781
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db   2783 cagcagttggtcagaacatctccttatcatgttctgaggctaccaggagtggctgattcg 2842

Qy   4782 gtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacctcca 4841
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2843 gtggtcaaacactgtgtgccctgccagctggttaatgctaatccttccagaatacctcca 2902

Qy   4842 ggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactgaggta 4901
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2903 ggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactgaggta 2962

Qy   4902 aagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcaggatgg 4961
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2963 aagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcaggatgg 3022

Qy   4962 gtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatactggag 5021
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3023 gtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatactggag 3082

Qy   5022 gaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagctttc 5081
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3083 gaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagctttc 3142

Qy   5082 gttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgcattgt 5141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3143 gttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgcattgt 3202

Qy   5142 gcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacc 5201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3203 gcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacc 3262

Qy   5202 cttac 5206
          |||||
Db   3263 cttac 3267
```

RESULT 9
AAC67021
ID AAC67021 standard; DNA; 6076 BP.
XX
AC AAC67021;
XX
DT 27-MAR-2001 (first entry)
XX
DE PERV env protein coding sequence SEQ ID NO: 21.
XX
KW Xenotransplantation; infectious agent; vaccine; ds.
XX
OS Porcine endogenous retrovirus.
XX
PN WO200071726-A1.
XX
PD 30-NOV-2000.
XX
PF 24-MAY-2000; 2000WO-US14296.
XX
PR 24-MAY-1999; 99US-0135631.
XX
PA (MAYO-) MAYO MEDICAL VENTURES.

XX
PI Federspiel MJ;
XX
DR WPI; 2001-032041/04.
XX
PT Inhibiting or preventing infectious agent transmission in mammalian
PT transplant recipients, by introducing recombinant DNA comprising DNA
PT encoding extracellular proteins of the agent into donor cells, such as
PT swine cells -
XX
PS Claim 16; Page 107-109; 144pp; English.
XX
CC The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX
SQ Sequence 6076 BP; 1613 A; 1512 C; 1609 G; 1342 T; 0 other;

Query Match 9.7%; Score 797; DB 22; Length 6076;
Best Local Similarity 98.7%; Pred. No. 1.8e-297;
Matches 2197; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```
Qy   2982 ttctgcctgagattacaccccactagccaaccacttttgccttcgaatggagagatcca 3041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3577 ttctgcctgagattacaccccactagccaaccacttttgccttcgaatggagagatcca 3636

Qy   3042 ggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactcc 3101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3637 ggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactcc 3696

Qy   3102 ccgaccatctttgacgaagccctacacagggacctggccaacttcaggatccaacaccct 3161
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   3697 ccgaccatctttgacgaagccctacacagagacctggccaacttcaggatccaacaccct 3756

Qy   3162 caggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggac 3221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3757 caggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggac 3816

Qy   3222 tgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctacagagcctct 3281
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   3817 tgcttagaaggcacgaaggcactactgctggaattgtctgacctaggctacagagcctct 3876

Qy   3282 gctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggc 3341
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   3877 gctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcgggac 3936

Qy   3342 gggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacc 3401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3937 gggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacc 3996

Qy   3402 acagccaaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccg 3461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3997 acagccaaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccg 4056

Qy   3462 gggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggggattctcc 3521
          |||||||||||||||||||||||||||||| |||||||||||||||||||| |||||||
Db   4057 gggtttgcgaccttagcagccccactctacccactaaccaaagaaaaaggggaattctcc 4116

Qy   3522 tgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacct 3581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4117 tgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacct 4176

Qy   3582 gctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaaggga 3641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4177 gctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaaggga 4236

Qy   3642 gtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctg 3701
          |||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   4237 gtagcccggggagttttaacccaaactctaggaccatggaggagacctgttgcctacctg 4296

Qy   3702 tcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagct 3761
          |||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||||
Db   4297 tcaaagaagctcgatcctgtagccagtggttggcccgtatgcctgaaggctatcgcagct 4356

Qy   3762 gtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaata 3821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4357 gtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaata 4416

Qy   3822 gccccccatgcattggagaacatcgttcggcagccccagaccgatggatgaccaacgcc 3881
          |||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4417 gcccaccatgcgttggagaacatcgttcggcagccccagaccgatggatgaccaacgcc 4476

Qy   3882 cgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgctccaccagcc 3941
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   4477 cgcatgacccactatcaaagcctgcttctcacagagagggtcacgttcgctccaccagcc 4536

Qy   3942 gctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgc 4001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4537 gctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgc 4596

Qy   4002 catcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgact 4061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4597 catcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgact 4656

Qy   4062 ggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatg 4121
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4657 ggagaagtgttaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatg 4716

Qy   4122 gctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctgccggaagga 4181
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4717 gctggggcggcggtggtggacgggacccgcacgatctgggccagcagcctgccggaagga 4776

Qy   4182 acttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaaggg 4241
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4777 acttcagcacaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaaggg 4836

Qy   4242 aaatccataaacatttatacggacagcaggtatgcctttgcgactgcacacgtacacggg 4301
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   4837 aaatccataaacatttatacagacagcaggtatgcctttgcgactgcacacgtacacggg 4896

Qy   4302 gccatctataaacaaaggggttgcttacctcagcagggagggaaataaagaacaaagag 4361
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4897 gccatctataagcaaaggggttgcttacctcagcagggagggaaataaagaacaaagag 4956

Qy   4362 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 4421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4957 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 5016

Qy   4422 cctggacatcagaaagccaaagatctcatatctagagggaaccagatggctgaccgggtt 4481
          |||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||
Db   5017 cctggacatcagaaagccaaagatcccatatccagagggaaccagatggctgaccgggtt 5076

Qy   4482 gccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaaagccccca 4541
          |||||||||||||||||||| |||||||||||||||||||| |||||||| ||||||||||
Db   5077 gccaagcaggcagcccagggtgttaaccttctgcctatgatagaaacacccaaagccccca 5136

Qy   4542 gaacccagacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttc 4601
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5137 gaacccggacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttc 5196

Qy   4602 tctgagactccggagggacctgctatacctcatatgggaaggaaatcctgccccacaaa 4661
          |||||||||||||| |||||||||||||||||| ||||||||||||||||||||||||||
Db   5197 tctgagactccggaagggacctgctatacctcagatgggaaggaaatcctgccccacaaa 5256

Qy   4662 gaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaaacacctg 4721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5257 gaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaaacacctg 5316

Qy   4722 cagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctgactcg 4781
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db   5317 cagcagttggtcagaacatctccttatcatgttctgaggctaccaggagtggctgattcg 5376

Qy   4782 gtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacctcca 4841
```

```
Db    5377 gtggtcaaacactgtgtgccctgccagctggttaatgctaatccttccagaatacctcca 5436

Qy    4842 ggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactgaggta 4901
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5437 ggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactgaggta 5496

Qy    4902 aagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcaggatgg 4961
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5497 aagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcaggatgg 5556

Qy    4962 gtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatactggag 5021
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5557 gtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatactggag 5616

Qy    5022 gaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagctttc 5081
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5617 gaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagctttc 5676

Qy    5082 gttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgcattgt 5141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5677 gttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgcattgt 5736

Qy    5142 gcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacc 5201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5737 gcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacc 5796

Qy    5202 cttac 5206
           |||||
Db    5797 cttac 5801
```

RESULT 10
AAC67019
ID AAC67019 standard; DNA; 7362 BP.
XX
AC AAC67019;
XX
DT 27-MAR-2001 (first entry)
XX
DE PERV env protein coding sequence SEQ ID NO: 19.
XX
KW Xenotransplantation; infectious agent; vaccine; ds.
XX
OS Porcine endogenous retrovirus.
XX
PN WO200071726-A1.
XX
PD 30-NOV-2000.
XX
PF 24-MAY-2000; 2000WO-US14296.
XX
PR 24-MAY-1999; 99US-0135631.
XX
PA (MAYO-) MAYO MEDICAL VENTURES.
XX
PI Federspiel MJ;
XX
DR WPI; 2001-032041/04.
XX
PT Inhibiting or preventing infectious agent transmission in mammalian
PT transplant recipients, by introducing recombinant DNA comprising DNA
PT encoding extracellular proteins of the agent into donor cells, such as
PT swine cells -
XX
PS Claim 16; Page 101-104; 144pp; English.
XX
CC The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX

SQ Sequence 7362 BP; 1997 A; 1821 C; 1881 G; 1663 T; 0 other;

Query Match 9.7%; Score 797; DB 22; Length 7362;
Best Local Similarity 98.7%; Pred. No. 1.8e-297;
Matches 2197; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```
Qy    2982 ttctgcctgagattacaccccactagccaaccacttttttgccttcgaatggagagatcca 3041
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3642 ttctgcctgagattacaccccactagccaaccacttttttgccttcgaatggagagatcca 3701

Qy    3042 ggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactcc 3101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3702 ggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactcc 3761

Qy    3102 ccgaccatctttgacgaagccctacacagggacctggccaacttcaggatccaacaccct 3161
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    3762 ccgaccatctttgacgaagccctacacagagacctggccaacttcaggatccaacaccct 3821

Qy    3162 caggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggac 3221
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3822 caggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggac 3881

Qy    3222 tgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctacagagcctct 3281
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    3882 tgcttagaaggcacgaaggcactactgctggaattgtctgacctaggctacagagcctct 3941

Qy    3282 gctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggc 3341
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    3942 gctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcgggac 4001

Qy    3342 gggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacc 3401
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4002 gggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacc 4061

Qy    3402 acagccaaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccg 3461
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4062 acagccaaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccg 4121

Qy    3462 gggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggggattctcc 3521
           |||||||||||||||||||||||||||||| |||||||||||||||||||| |||||||
Db    4122 gggtttgcgaccttagcagccccactctacccactaaccaaagaaaaaggggaattctcc 4181

Qy    3522 tgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacct 3581
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4182 tgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacct 4241

Qy    3582 gctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaaggga 3641
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4242 gctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaaggga 4301

Qy    3642 gtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctg 3701
           |||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    4302 gtagcccggggagttttaacccaaactctaggaccatggaggagacctgttgcctacctg 4361

Qy    3702 tcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagct 3761
           |||||||||||| |||||||||||||||||||||||||| ||||||||||||||||||||
Db    4362 tcaaagaagctcgatcctgtagccagtggttggcccgtatgcctgaaggctatcgcagct 4421

Qy    3762 gtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaata 3821
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4422 gtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaata 4481

Qy    3822 gccccccatgcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgcc 3881
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    4482 gccccccatgcgttggagaacatcgttcggcagcccccagaccgatggatgaccaacgcc 4541

Qy    3882 cgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgctccaccagcc 3941
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    4542 cgcatgacccactatcaaagcctgcttctcacagagagggtcacgttcgctccaccagcc 4601

Qy    3942 gctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgc 4001
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   4602 gctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgc 4661

Qy   4002 catcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgact 4061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4662 catcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgact 4721

Qy   4062 ggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatg 4121
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4722 ggagaagtgttaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatg 4781

Qy   4122 gctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctgccggaagga 4181
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4782 gctggggcggcggtggtggacgggacccgcacgatctgggccagcagcctgccggaagga 4841

Qy   4182 acttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaaggg 4241
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4842 acttcagcacaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaaggg 4901

Qy   4242 aaatccataaacatttatacggacagcaggtatgcctttgcgactgcacacgtacacggg 4301
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   4902 aaatccataaacatttatacagacagcaggtatgcctttgcgactgcacacgtacacggg 4961

Qy   4302 gccatctataaacaaaggggttgcttacctcagcagggagggaaataaagaacaaagag 4361
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4962 gccatctataagcaaaggggttgcttacctcagcagggagggaaataaagaacaaagag 5021

Qy   4362 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 4421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5022 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 5081

Qy   4422 cctggacatcagaaagccaaagatctcatatctagagggaaccagatggctgaccgggtt 4481
          |||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||||||
Db   5082 cctggacatcagaaagccaaagatcccatatccagagggaaccagatggctgaccgggtt 5141

Qy   4482 gccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaaagcccca 4541
          |||||||||||||||||| |||||||||||||||||| |||||||| ||||||||||||||
Db   5142 gccaagcaggcagcccagggtgttaaccttctgcctatgatagaaacacccaaagcccca 5201

Qy   4542 gaacccagacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttc 4601
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5202 gaacccggacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttc 5261

Qy   4602 tctgagactccggagggacctgctatacctcatatgggaaggaaatcctgccccacaaa 4661
          |||||||||||||| |||||||||||||||||||| ||||||||||||||||||||||||||
Db   5262 tctgagactccggaagggacctgctatacctcagatgggaaggaaatcctgccccacaaa 5321

Qy   4662 gaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaaacacctg 4721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5322 gaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaaacacctg 5381

Qy   4722 cagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctgactcg 4781
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||| |||
Db   5382 cagcagttggtcagaacatctccttatcatgttctgaggctaccaggagtggctgattcg 5441

Qy   4782 gtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacctcca 4841
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5442 gtggtcaaacactgtgtgccctgccagctggttaatgctaatccttccagaatacctcca 5501

Qy   4842 ggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactgaggta 4901
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5502 ggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactgaggta 5561

Qy   4902 aagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcaggatgg 4961
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5562 aagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcaggatgg 5621

Qy   4962 gtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatactggag 5021
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   5622 gtagaggcttatcctactaagaaagagacttcaaccgtggtggctaaaaaaatactggag 5681

Qy   5022 gaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagctttc 5081
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5682 gaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagctttc 5741

Qy   5082 gttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgcattgt 5141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5742 gttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgcattgt 5801

Qy   5142 gcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacc 5201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5802 gcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacc 5861

Qy   5202 cttac 5206
          |||||
Db   5862 cttac 5866
```

RESULT 11
AAC67032
ID AAC67032 standard; DNA; 6076 BP.
XX
AC AAC67032;
XX
DT 27-MAR-2001 (first entry)
XX
DE PERV env protein coding sequence SEQ ID NO: 32.
XX
KW Xenotransplantation; infectious agent; vaccine; ds.
XX
OS Porcine endogenous retrovirus.
XX
PN WO200071726-A1.
XX
PD 30-NOV-2000.
XX
PF 24-MAY-2000; 2000WO-US14296.
XX
PR 24-MAY-1999; 99US-0135631.
XX
PA (MAYO-) MAYO MEDICAL VENTURES.
XX
PI Federspiel MJ;
XX
DR WPI; 2001-032041/04.
XX
PT Inhibiting or preventing infectious agent transmission in mammalian
PT transplant recipients, by introducing recombinant DNA comprising DNA
PT encoding extracellular proteins of the agent into donor cells, such as
PT swine cells -
XX
PS Claim 16; Page 117-119; 144pp; English.
XX
CC The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX
SQ Sequence 6076 BP; 1619 A; 1505 C; 1598 G; 1354 T; 0 other;

Query Match 9.1%; Score 746; DB 22; Length 6076;
Best Local Similarity 98.7%; Pred. No. 7.7e-278;
Matches 2196; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
Qy   2982 ttctgcctgagattacaccccactagccaaccacttttgccttcgaatggagagatcca 3041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3577 ttctgcctgagattacaccccactagccaaccacttttgccttcgaatggagagatcca 3636

Qy   3042 ggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactcc 3101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3637 ggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactcc 3696

Qy   3102 ccgaccatctttgacgaagccctacacagggacctggccaacttcaggatccaacaccct 3161
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   3697 ccgaccatctttgacgaagccctacacagagacctggccaacttcaggatccaacaccct 3756
```

```
Qy   3162 caggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggac 3221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3757 caggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggac 3816

Qy   3222 tgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctacagagcctct 3281
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3817 tgcttagaaggcacgaaggcactactgctggaattgtctgacctaggctacagagcctct 3876

Qy   3282 gctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggc 3341
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   3877 gctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcgggac 3936

Qy   3342 gggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacc 3401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3937 gggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacc 3996

Qy   3402 acagccaaacaagtgagagagttttggggacagctggattttgcagactgtggatcccg 3461
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3997 acagccaaacaagtgagagaattttggggacagctggattttgcagactgtggatcccg 4056

Qy   3462 gggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggggattctcc 3521
          |||||||||||||||||||||||||||||||| |||||||||||||||||| |||||||
Db   4057 gggtttgcgaccttagcagccccactctacccactaaccaaagaaaaaggggaattctcc 4116

Qy   3522 tgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacct 3581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4117 tgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacct 4176

Qy   3582 gctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaaggga 3641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4177 gctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaaggga 4236

Qy   3642 gtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctg 3701
          |||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   4237 gtagcccggggagttttaacccaaactctaggaccatggaggagacctgttgcctacctg 4296

Qy   3702 tcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagct 3761
          |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
Db   4297 tcaaagaagctcgatcctgtagccagtggttggcccgtatgcctgaaggctatcgcagct 4356

Qy   3762 gtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaata 3821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4357 gtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaata 4416

Qy   3822 gccccccatgcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgcc 3881
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4417 gccccccatgcgttggagaacatcgttcggcagcccccagaccgatggatgaccaacgcc 4476

Qy   3882 cgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgctccaccagcc 3941
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   4477 cgcatgacccactatcaaagcctgcttctcacagagagggtcacgttcgctccaccagcc 4536

Qy   3942 gctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgc 4001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4537 gctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgc 4596

Qy   4002 catcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgact 4061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4597 catcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgact 4656

Qy   4062 ggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatg 4121
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4657 ggagaagtgttaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatg 4716

Qy   4122 gctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctgccggaagga 4181
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4717 gctggggcggcggtggtggacgggacccgcacgatctgggccagcagcctgccggaagga 4776

Qy   4182 acttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaaggg 4241
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4777 acttcagcacaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaaggg 4836
```

```
Qy   4242 aaatccataaacatttatacggacagcaggtatgcctttgcgactgcacacgtacacggg 4301
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   4837 aaatccataaacatttatacagacagcaggtatgcctttgcgactgcacacgtacacggg 4896

Qy   4302 gccatctataaacaaaggggggttgcttacctcagcagggagggaaataaagaacaaagag 4361
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4897 gccatctataagcaaaggggggttgcttacctcagcagggagggaaataaagaacaaagag 4956

Qy   4362 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 4421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4957 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 5016

Qy   4422 cctggacatcagaaagccaaagatctcatatctagagggaaccagatggctgaccgggtt 4481
          |||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||||
Db   5017 cctggacatcagaaagccaaagatcccatatccagagggaaccagatggctgaccgggtt 5076

Qy   4482 gccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaaagcccca 4541
          |||||||||||||||||||| |||||||||||||||||| |||||||| ||||||||||||
Db   5077 gccaagcaggcagcccagggtgttaaccttctgcctatgatagaaacacccaaagcccca 5136

Qy   4542 gaacccagacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttc 4601
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5137 gaacccggacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttc 5196

Qy   4602 tctgagactccggaggggacctgctatacctcatatgggaaggaaatcctgccccacaaa 4661
          |||||||||||||| |||||||||||||||||| ||||||||||||||||||||||||||
Db   5197 tctgagactccggaagggacctgctatacctcagatgggaaggaaatcctgccccacaaa 5256

Qy   4662 gaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaaacacctg 4721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5257 gaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaaacacctg 5316

Qy   4722 cagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctgactcg 4781
          |||||||||||||||||||| |||||||||||||||||||||||||||||||| |||||
Db   5317 cagcagttggtcagaacatctccttatcatgttctgaggctaccaggagtggctgattcg 5376

Qy   4782 gtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacctcca 4841
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5377 gtggtcaaacactgtgtgccctgccagctggttaatgctaatccttccagaatacctcca 5436

Qy   4842 ggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactgaggta 4901
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5437 ggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactgaggta 5496

Qy   4902 aagccggctaaatacggaaacaaatatctattggtttttgtagacacctttttcaggatgg 4961
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5497 aagccggctaaatacggaaacaaatatctattggtttttgtagacacctttttcaggatgg 5556

Qy   4962 gtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatactggag 5021
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5557 gtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatactggag 5616

Qy   5022 gaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagctttc 5081
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5617 gaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagctttc 5676

Qy   5082 gttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgcattgt 5141
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   5677 gttgcccaggtaagtcagggactggccaagatattggggattaattggaaactgcattgt 5736

Qy   5142 gcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacc 5201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5737 gcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacc 5796

Qy   5202 cttac 5206
          |||||
Db   5797 cttac 5801
```

RESULT 12
AAT74884
ID AAT74884 standard; cDNA; 7892 BP.
XX

```
AC   AAT74884;
XX
DT   09-FEB-1998  (first entry)
XX
DE   Miniature swine retrovirus cDNA.
XX
KW   Retrovirus; porcine; GAG protein; POL protein; ENV protein;
KW   xenotransplantation; infectious; provirus; organ transplant; donor;
KW   activated virus; PCR; ss.
XX
OS   Porcine retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             585..2159
FT                   /*tag= a
FT   mat_peptide     585..2156
FT                   /*tag= b
FT                   /note= "putative GAG protein"
FT   CDS             2307..5744
FT                   /*tag= c
FT   mat_peptide     2307..5741
FT                   /*tag= d
FT                   /note= "putative POL protein"
FT   CDS             5620..7536
FT                   /*tag= e
FT   mat_peptide     5620..7533
FT                   /*tag= f
FT                   /note= "putative ENV protein"
XX
PN   WO9721836-A1.
XX
PD   19-JUN-1997.
XX
PF   13-DEC-1996;   96WO-US19680.
XX
PR   14-DEC-1995;   95US-0572645.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Fishman JA;
XX
DR   WPI; 1997-332804/30.
DR   P-PSDB; AAW32096-W32098.
XX
PT   New nucleic acid from porcine retro:viruses - used for detecting
PT   viruses in transplant or other tissue and for assessing risk of
PT   transmitting infection to graft recipient
XX
PS   Claim 22; Fig 3; 128pp; English.
XX
CC   This cDNA sequence represents a porcine retrovirus from miniature swine
CC   containing the putative coding regions for viral GAG, POL and ENV
CC   proteins. This sequence and PCR fragments generated from it
CC   (see AAT74812-T74882) can be used to screen organs for the presence of
CC   porcine retroviruses prior to xenotransplantation. Transplantation can
CC   increase the likelihood of retroviral activation if intact and
CC   infectious proviruses are present. The porcine retroviral sequence can be
CC   used to generate probes to determine the level (e.g. copy number) of
CC   intact (i.e. potentially replicating) porcine provirus sequences in a
CC   strain of xenograft transplantation donors. It can be used to detect
CC   mutations, genetic lesions or viral recombinants and also to determine
CC   the histological localisation of activated retrovirus. Using Polymerase
CC   Chain Reaction DNA Quantitation (PDQ) on blood mononuclear cells,
CC   infectivity titration and susceptibility testing can be performed.
CC   Ultimately animal donors without intact porcine retroviral sequences or a
CC   lower copy number of viral elements could be selected.
XX
SQ   Sequence 7892 BP; 2191 A; 1915 C; 1980 G; 1806 T; 0 other;

  Query Match             8.6%;  Score 704;  DB 18;  Length 7892;
  Best Local Similarity   98.7%;  Pred. No. 1.1e-261;
  Matches 2204;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy   2973 gatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcgaatgg 3032
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2970 gatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcgaatgg 3029

Qy   3033 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 3092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3030 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 3089

Qy   3093 aagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatc 3152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3090 aagaactccccgaccatctttgacgaagccctacacagagacctggccaacttcaggatc 3149

Qy   3153 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 3212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3150 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 3209

Qy   3213 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 3272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3210 aaacaggactgcttagaaggcacgaaggcactactgctggaattgtctgacctaggctac 3269

Qy   3273 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 3332
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3270 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 3329

Qy   3333 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 3392
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3330 ttgcgggacgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 3389

Qy   3393 gccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagactg 3452
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   3390 gccccaaccacagccaaacaaatgagagagtttttggggacagctggattttgcagactg 3449

Qy   3453 tggatcccgggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 3512
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3450 tggatcccgggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 3509

Qy   3513 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 3572
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3510 gaattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 3569

Qy   3573 agcgcacctgctctggccctccctgacgtaactaaaccctttacccttatgtggatgag 3632
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3570 agcgcacctgctctggccctccctgacgtaactaaaccctttacccttatgtggatgag 3629

Qy   3633 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 3692
           |||||||||||||||| |||||||||||||||||||||||||||||||| ||||||||
Db   3630 cgtaagggagtagcccggggagttttaacccaaaccctaggaccatggagaagacctgtc 3689

Qy   3693 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 3752
           |||||||||||||||||| |||||||||||||||||||||||||| |||| |||||||||
Db   3690 gcctacctgtcaaagaagctcgatcctgtagccagtggttggcccatatgcctgaaggct 3749

Qy   3753 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatata 3812
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3750 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatata 3809

Qy   3813 actgtaatagccccccatgcattggagaacatcgttcggcagcccccagaccgatggatg 3872
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3810 actgtaatagccccccatgcattggagaacatcgttcggcagcccccagaccgatggatg 3869

Qy   3873 accaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgct 3932
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   3870 accaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcacgttcgct 3929

Qy   3933 ccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgact 3992
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3930 ccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgact 3989

Qy   3993 catgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacata 4052
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3990 catgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacata 4049
```

```
Qy   4053 ccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggt 4112
          ||||||
```

```
Db   3090 aagaactccccgaccatctttgacgaagccctacacagagacctggccaacttcaggatc 3149

Qy   3153 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 3212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3150 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 3209

Qy   3213 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 3272
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3210 aaacaggactgcttagaaggcacgaaggcactactgctggaattgtctgacctaggctac 3269

Qy   3273 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 3332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3270 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 3329

Qy   3333 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 3392
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3330 ttgcgggacgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 3389

Qy   3393 gccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagactg 3452
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3390 gccccaaccacagccaaacaaatgagagagtttttggggacagctggattttgcagactg 3449

Qy   3453 tggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 3512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3450 tggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 3509

Qy   3513 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 3572
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3510 gaattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 3569

Qy   3573 agcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgag 3632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3570 agcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgag 3629

Qy   3633 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 3692
          |||||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Db   3630 cgtaagggagtagcccggggagttttaacccaaaccctaggaccatggagaagacctgtc 3689

Qy   3693 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 3752
          |||||||||||||||||| |||||||||||||||||||||||||| |||| |||||||||
Db   3690 gcctacctgtcaaagaagctcgatcctgtagccagtggttggcccatatgcctgaaggct 3749

Qy   3753 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatata 3812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3750 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatata 3809

Qy   3813 actgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatggatg 3872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3810 actgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatggatg 3869

Qy   3873 accaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgct 3932
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   3870 accaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcacgttcgct 3929

Qy   3933 ccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgact 3992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3930 ccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgact 3989

Qy   3993 catgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacata 4052
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3990 catgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacata 4049

Qy   4053 ccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggt 4112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4050 ccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggt 4109

Qy   4113 aagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctg 4172
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   4110 aagaggatggctggggcggcggtggtggacgggacccgcacgatctgggccagcagcctg 4169

Qy   4173 ccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctg 4232
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   4170 ccggaaggaacttcagcacaaaaggctgagctcatggccctcacgcaagctttgcggctg 4229
```

```
Qy   4233 gccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcacac 4292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4230 gccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcacac 4289

Qy   4293 gtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaag 4352
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4290 gtacatggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaag 4349

Qy   4353 aacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctatt 4412
          |||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||
Db   4350 aacaaagaggaaattctaagcctattagaagccgtacatttaccaaaaaggctagctatt 4409

Qy   4413 atacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggct 4472
          |||||||||||||||||||||||||| |||||||||||||| ||||| ||||||||||||
Db   4410 atacactgtcctggacatcagaaagctaaagatctcatatccagaggaaaccagatggct 4469

Qy   4473 gaccgggttgccaagcaggcagcccaggctgttaacccttctgcctataatagaaacgccc 4532
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||| ||||
Db   4470 gaccgggttgccaagcaggcagcccagggtgttaacccttctgcctataatagaaatgccc 4529

Qy   4533 aaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagata 4592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4530 aaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagata 4589

Qy   4593 gaccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaatcctg 4652
          |||||||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||
Db   4590 gaccagttctctgagactccggaagggacctgctatacctcagatgggaaggaaatcctg 4649

Qy   4653 ccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaact 4712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4650 ccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaact 4709

Qy   4713 aaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtg 4772
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4710 aaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtg 4769

Qy   4773 gctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccaga 4832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4770 gctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccaga 4829

Qy   4833 atacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttc 4892
          || |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4830 atgcctccagggaagagactaaggggaagccacccaggcgctcactgggaagtggacttc 4889

Qy   4893 actgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacaccttt 4952
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   4890 actgaggtaaagccggctaaatacggaaacaaatacctattggtttttgtagacaccttt 4949

Qy   4953 tcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaa 5012
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   4950 tcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaaaaaa 5009

Qy   5013 atactggaggaaatttttccaagatttggaatacctaaggtaataggtcagacaatggt 5072
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5010 atactggaagaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggt 5069

Qy   5073 ccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaa 5132
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5070 ccagcttttgttgcccaggtaagtcagggactggccaagatattggggattgattggaaa 5129

Qy   5133 ctgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccatt 5192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5130 ctgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccatt 5189

Qy   5193 aaagagacccttac 5206
          ||||||||||||||
Db   5190 aaagagacccttac 5203
```

RESULT 14
AAC67022
ID AAC67022 standard; DNA; 4918 BP.

XX
AC AAC67022;
XX
DT 27-MAR-2001 (first entry)
XX
DE PERV env protein coding sequence SEQ ID NO: 22.
XX
KW Xenotransplantation; infectious agent; vaccine; ds.
XX
OS Porcine endogenous retrovirus.
XX
PN WO200071726-A1.
XX
PD 30-NOV-2000.
XX
PF 24-MAY-2000; 2000WO-US14296.
XX
PR 24-MAY-1999; 99US-0135631.
XX
PA (MAYO-) MAYO MEDICAL VENTURES.
XX
PI Federspiel MJ;
XX
DR WPI; 2001-032041/04.
XX
PT Inhibiting or preventing infectious agent transmission in mammalian
PT transplant recipients, by introducing recombinant DNA comprising DNA
PT encoding extracellular proteins of the agent into donor cells, such as
PT swine cells -
XX
PS Claim 16; Page 109-111; 144pp; English.
XX
CC The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX
SQ Sequence 4918 BP; 1382 A; 1173 C; 1183 G; 1180 T; 0 other;

Query Match 6.9%; Score 570; DB 22; Length 4918;
Best Local Similarity 98.7%; Pred. No. 4.4e-210;
Matches 1670; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy   3515 attctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgag 3574
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     79 attctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgag 138

Qy   3575 cgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcg 3634
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    139 cgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcg 198

Qy   3635 taagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgttgc 3694
          |||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||||
Db    199 taagggagtagcccggggagttttaacccaaactctaggaccatggaggagacctgttgc 258

Qy   3695 ctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctat 3754
          |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||
Db    259 ctacctgtcaaagaagctcgatcctgtagccagtggttggcccgtatgcctgaaggctat 318

Qy   3755 cgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatataac 3814
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 cgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatataac 378

Qy   3815 tgtaatagccccccatgcattggagaacatcgttcggcagcccccagaccgatggatgac 3874
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    379 tgtaatagccccccatgcgttggagaacatcgttcggcagcccccagaccgatggatgac 438

Qy   3875 caacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgctcc 3934
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    439 caacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcacgttcgctcc 498
```

```
Qy   3935 accagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactca 3994
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 accagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactca 558

Qy   3995 tgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacatacc 4054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    559 tgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacatacc 618

Qy   4055 gctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggtaa 4114
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    619 gctgactggagaagtgttaacctggttcactgacggaagcagctatgtggtggaaggtaa 678

Qy   4115 gaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctgcc 4174
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    679 gaggatggctggggcggcggtggtggacgggacccgcacgatctgggccagcagcctgcc 738

Qy   4175 ggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggc 4234
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    739 ggaaggaacttcagcacaaaaggctgagctcatggccctcacgcaagctttgcggctggc 798

Qy   4235 cgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcacacgt 4294
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    799 cgaagggaaatccataaacatttatacagacagcaggtatgcctttgcgactgcacacgt 858

Qy   4295 acacggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaagaa 4354
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    859 acacggggccatctataagcaaaggggttgcttacctcagcagggagggaaataaagaa 918

Qy   4355 caaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctattat 4414
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    919 caaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctattat 978

Qy   4415 acactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggctga 4474
          |||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||
Db    979 acactgtcctggacatcagaaagccaaagatcccatatccagagggaaccagatggctga 1038

Qy   4475 ccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaa 4534
          |||||||||||||||||||||||||| |||||||||||||||| |||||||| |||||
Db   1039 ccgggttgccaagcaggcagcccagggtgttaaccttctgcctatgatagaaacacccaa 1098

Qy   4535 agccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagataga 4594
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1099 agccccagaacccggacgacagtacaccctagaagactggcaagagataaaaaagataga 1158

Qy   4595 ccagttctctgagactccggagggggacctgctatacctcatatgggaaggaaatcctgcc 4654
          |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||
Db   1159 ccagttctctgagactccggaagggacctgctatacctcagatgggaaggaaatcctgcc 1218

Qy   4655 ccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaa 4714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1219 ccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaa 1278

Qy   4715 acacctgcagcagttggtcagaacatcccctatcatgttctgaggctaccaggagtggc 4774
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   1279 acacctgcagcagttggtcagaacatctccttatcatgttctgaggctaccaggagtggc 1338

Qy   4775 tgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaat 4834
          ||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1339 tgattcggtggtcaaacactgtgtgccctgccagctggttaatgctaatccttccagaat 1398

Qy   4835 acctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcac 4894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1399 acctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcac 1458

Qy   4895 tgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctttc 4954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1459 tgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctttc 1518

Qy   4955 aggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaat 5014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1519 aggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaat 1578

Qy   5015 actggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtcc 5074
```

```
Db   1579 actggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtcc 1638

Qy   5075 agctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaact 5134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1639 agctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaact 1698

Qy   5135 gcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaa 5194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1699 gcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaa 1758

Qy   5195 agagacccttac 5206
          ||||||||||||
Db   1759 agagacccttac 1770


RESULT 15
AAC67023
ID   AAC67023 standard; DNA; 7873 BP.
XX
AC   AAC67023;
XX
DT   27-MAR-2001  (first entry)
XX
DE   PERV env protein coding sequence SEQ ID NO: 23.
XX
KW   Xenotransplantation; infectious agent; vaccine; ds.
XX
OS   Porcine endogenous retrovirus.
XX
PN   WO200071726-A1.
XX
PD   30-NOV-2000.
XX
PF   24-MAY-2000; 2000WO-US14296.
XX
PR   24-MAY-1999;   99US-0135631.
XX
PA   (MAYO-) MAYO MEDICAL VENTURES.
XX
PI   Federspiel MJ;
XX
DR   WPI; 2001-032041/04.
XX
PT   Inhibiting or preventing infectious agent transmission in mammalian
PT   transplant recipients, by introducing recombinant DNA comprising DNA
PT   encoding extracellular proteins of the agent into donor cells, such as
PT   swine cells -
XX
PS   Claim 16; Page 112-115; 144pp; English.
XX
CC   The present invention provides a method to prevent the transmission of
CC   infectious agents during xenotransplantation. This involves introducing
CC   to donor swine cells a recombinant DNA encoding a peptide fragment from
CC   the infectious agent, and then introducing these cells into the
CC   transplant recipient.
XX
SQ   Sequence 7873 BP; 2200 A; 1913 C; 1962 G; 1798 T; 0 other;


  Query Match             6.0%;  Score 496;  DB 22;  Length 7873;
  Best Local Similarity   98.7%;  Pred. No. 1.3e-181;
  Matches 1446;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy   2982 ttctgcctgagattacaccccactagccaaccacttttgccttcgaatggagagatcca 3041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2456 ttctgcctgagattacaccccactagccaaccacttttgccttcgaatggagagatcca 2515

Qy   3042 ggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactcc 3101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2516 ggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactcc 2575

Qy   3102 ccgaccatctttgacgaagccctacacagggacctggccaacttcaggatccaacaccct 3161
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2576 ccgaccatctttgacgaagccctacacagagacctggccaacttcaggatccaacaccct 2635

Qy   3162 caggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggac 3221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2636 caggtgaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggac 2695

Qy   3222 tgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctacagagcctct 3281
          |||||||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2696 tgcttagaaagcacgaaggcactactgctggaattgtctgacctaggctacagagcctct 2755

Qy   3282 gctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggc 3341
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2756 gctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcgggac 2815

Qy   3342 gggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacc 3401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2816 gggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacc 2875

Qy   3402 acagccaaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccg 3461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2876 acagccaaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccg 2935

Qy   3462 gggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggggattctcc 3521
          |||||||||||||||||||||||||||||||| ||||||||||||||||||| |||||||
Db   2936 gggtttgcgaccttagcagccccactctacccactaaccaaagaaaaaggggaattctcc 2995

Qy   3522 tgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacct 3581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2996 tgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacct 3055

Qy   3582 gctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaaggga 3641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3056 gctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaaggga 3115

Qy   3642 gtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctg 3701
          |||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   3116 gtagcccggggagttttaacccaaactctaggaccatggaggagacctgttgcctacctg 3175

Qy   3702 tcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagct 3761
          |||||||||||| |||||||||||||||||||||||||| |||||||||||||||||||||
Db   3176 tcaaagaagctcgatcctgtagccagtggttggcccgtatgcctgaaggctatcgcagct 3235

Qy   3762 gtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaata 3821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3236 gtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaata 3295

Qy   3822 gccccccatgcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgcc 3881
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3296 gccccccatgcgttggagaacatcgttcggcagcccccagaccgatggatgaccaacgcc 3355

Qy   3882 cgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgctccaccagcc 3941
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   3356 cgcatgacccactatcaaagcctgcttctcacagagagggtcacgttcgctccaccagcc 3415

Qy   3942 gctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgc 4001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3416 gctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgc 3475

Qy   4002 catcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgact 4061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3476 catcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgact 3535

Qy   4062 ggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatg 4121
          ||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   3536 ggagaagtgttaacctggttcactgacggaagcagctatgtagtggaaggtaagaggatg 3595

Qy   4122 gctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctgccggaagga 4181
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3596 gctggggcggcggtggtggacgggacccgcacgatctgggccagcagcctgccggaagga 3655

Qy   4182 acttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaaggg 4241
          | |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    3656 agttcagcacaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaaggg 3715

Qy    4242 aaatccataaacatttatacggacagcaggtatgcctttgcgactgcacacgtacacggg 4301
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    3716 aaatccataaacatttatacagacagcaggtatgcctttgcgactgcacacgtacacggg 3775

Qy    4302 gccatctataaacaaaggggttgcttacctcagcagggagggaaataaagaacaaagag 4361
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    3776 gccatctataagcaaaggggttgcttacctcagcagggagggaaataaagaacaaagag 3835

Qy    4362 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 4421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3836 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 3895

Qy    4422 cctggacatcagaaagccaaagatc 4446
           |||||||||||||||||||||||||
Db    3896 cctggacatcagaaagccaaagatc 3920
```

Search completed: February 24, 2002, 07:42:18
Job time: 16843 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 03:05:29 ; Search time 624.99 Seconds
(without alignments)
24.691 Million cell updates/sec

Title: US-09-171-553B-14
Perfect score: 18
Sequence: 1 ccacagtcgtacaccacg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	18	100.0	633	22	AAC67007	PERV-MSL LTR SEQ I
c	2	18	100.0	704	22	AAC67006	IamA1 LTR SEQ ID N
c	3	18	100.0	3482	20	AAV82749	Pig endogenous ret
c	4	18	100.0	4918	22	AAC67022	PERV env protein c
c	5	18	100.0	6076	22	AAC67021	PERV env protein c
c	6	18	100.0	6076	22	AAC67032	PERV env protein c
c	7	18	100.0	7333	22	AAF77726	Defective retrovir
c	8	18	100.0	7362	22	AAC67019	PERV env protein c
c	9	18	100.0	7393	18	AAT74883	Porcine retrovirus
c	10	18	100.0	7873	22	AAC67023	PERV env protein c
c	11	18	100.0	7892	18	AAT74884	Miniature swine re
c	12	18	100.0	8060	18	AAT74811	Porcine retrovirus
c	13	18	100.0	8060	22	AAF77725	Tsukuba-1 cDNA. P
c	14	18	100.0	8132	22	AAF77727	Nucleotide sequenc
c	15	18	100.0	8196	18	AAV09699	Porcine retrovirus
c	16	18	100.0	8209	18	AAV09700	Porcine retrovirus
	17	17	94.4	20	18	AAT74869	Porcine retrovirus
	18	17	94.4	20	22	AAF77785	PCR primer #58. U
c	19	17	94.4	28	20	AAV82760	PCR primer PL148 u
	20	15.4	85.6	6511	16	AAQ95493	Human Cdn-2 DNA.
	21	14.8	82.2	285	22	AAH50831	Human tumour assoc
c	22	14.8	82.2	752	22	AAH04528	Human cDNA clone (
	23	14.8	82.2	935	21	AAF09918	Fusarium venenatum
c	24	14.8	82.2	4299	19	AAV03310	Pyrolobus fumarius
	25	14.4	80.0	3178	20	AAX99865	Human secreted pro
c	26	14.4	80.0	15894	19	AAV18265	Measles virus 1977
c	27	14.4	80.0	15894	20	AAZ22902	Nucleotide sequenc
c	28	14.4	80.0	32249	22	AAI63718	Human kidney relat
c	29	13.8	76.7	130	21	AAC14197	Human secreted pro
	30	13.8	76.7	131	21	AAC29646	Human secreted pro
c	31	13.8	76.7	155	21	AAC13905	Human secreted pro
c	32	13.8	76.7	178	21	AAC13629	Human secreted pro
c	33	13.8	76.7	198	21	AAC23537	Human secreted pro
	34	13.8	76.7	300	20	AAZ13103	Human gene express
	35	13.8	76.7	300	20	AAX98595	Human cancer cell
c	36	13.8	76.7	318	21	AAC28819	Human secreted pro
c	37	13.8	76.7	344	21	AAC32109	Human secreted pro
	38	13.8	76.7	391	22	AAF66266	Novel human polynu
	39	13.8	76.7	417	21	AAC24025	Human secreted pro
	40	13.8	76.7	432	22	AAH13493	Human cDNA clone (
	41	13.8	76.7	530	21	AAZ49231	Human hydrolase ho
c	42	13.8	76.7	537	22	AAH10051	Human cDNA clone (
	43	13.8	76.7	583	22	AAH10021	Human cDNA clone (
	44	13.8	76.7	585	22	AAH10314	Human cDNA clone (
c	45	13.8	76.7	739	22	AAH08480	Human cDNA clone (

ALIGNMENTS

RESULT 1
AAC67007/c
ID AAC67007 standard; DNA; 633 BP.
XX
AC AAC67007;
XX
DT 27-MAR-2001 (first entry)
XX
DE PERV-MSL LTR SEQ ID NO: 7.
XX
KW Xenotransplantation; infectious agent; vaccine; ds.
XX
OS Porcine endogenous retrovirus.
XX
PN WO200071726-A1.
XX
PD 30-NOV-2000.
XX
PF 24-MAY-2000; 2000WO-US14296.
XX
PR 24-MAY-1999; 99US-0135631.
XX
PA (MAYO-) MAYO MEDICAL VENTURES.
XX
PI Federspiel MJ;
XX
DR WPI; 2001-032041/04.
XX
PT Inhibiting or preventing infectious agent transmission in mammalian
PT transplant recipients, by introducing recombinant DNA comprising DNA
PT encoding extracellular proteins of the agent into donor cells, such as
PT swine cells -
XX
PS Disclosure; Page 98; 144pp; English.

```
XX
CC   The present invention provides a method to prevent the transmission of
CC   infectious agents during xenotransplantation. This involves introducing
CC   to donor swine cells a recombinant DNA encoding a peptide fragment from
CC   the infectious agent, and then introducing these cells into the
CC   transplant recipient.
XX
SQ   Sequence 633 BP; 170 A; 125 C; 147 G; 191 T; 0 other;


  Query Match            100.0%;  Score 18;  DB 22;  Length 633;
  Best Local Similarity  100.0%;  Pred. No. 2.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          ||||||||||||||||||
Db    497 CCACAGTCGTACACCACG 480


RESULT   2
AAC67006/c
ID   AAC67006 standard; DNA; 704 BP.
XX
AC   AAC67006;
XX
DT   27-MAR-2001  (first entry)
XX
DE   IamA1 LTR SEQ ID NO: 6.
XX
KW   Xenotransplantation; infectious agent; vaccine; ds.
XX
OS   Unidentiifed.
XX
PN   WO200071726-A1.
XX
PD   30-NOV-2000.
XX
PF   24-MAY-2000; 2000WO-US14296.
XX
PR   24-MAY-1999;   99US-0135631.
XX
PA   (MAYO-) MAYO MEDICAL VENTURES.
XX
PI   Federspiel MJ;
XX
DR   WPI; 2001-032041/04.
XX
PT   Inhibiting or preventing infectious agent transmission in mammalian
PT   transplant recipients, by introducing recombinant DNA comprising DNA
PT   encoding extracellular proteins of the agent into donor cells, such as
PT   swine cells -
XX
PS   Disclosure; Page 97-98; 144pp; English.
XX
CC   The present invention provides a method to prevent the transmission of
CC   infectious agents during xenotransplantation. This involves introducing
CC   to donor swine cells a recombinant DNA encoding a peptide fragment from
CC   the infectious agent, and then introducing these cells into the
CC   transplant recipient.
XX
SQ   Sequence 704 BP; 187 A; 162 C; 158 G; 197 T; 0 other;


  Query Match            100.0%;  Score 18;  DB 22;  Length 704;
  Best Local Similarity  100.0%;  Pred. No. 2.5;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          ||||||||||||||||||
Db    568 CCACAGTCGTACACCACG 551


RESULT   3
AAV82749/c
ID   AAV82749 standard; DNA; 3482 BP.
XX
AC   AAV82749;
XX
DT   25-FEB-1999  (first entry)
XX
DE   Pig endogenous retrovirus (PERV)-B envelope (env) gene region.
XX
KW   Pig endogenous retrovirus; PERV-A; envelope protein; PERV-B; subtype;
KW   probe; primer; detection; retrovirus; human tissue; xenotransplant;
KW   primary porcine tissue; human cell line; porcine cell line; ss.
XX
OS   Pig endogenous retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             911..2884
FT                   /*tag=  a
FT                   /product= envelope_protein
XX
PN   WO9853104-A2.
XX
PD   26-NOV-1998.
XX
PF   18-MAY-1998;   98WO-GB01428.
XX
PR   16-MAY-1997;   97GB-0010154.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Stoye JP,  Weiss RA;
XX
DR   WPI; 1999-045324/04.
DR   P-PSDB; AAW85452.
XX
PT   Newly isolated nucleic acid probe capable of hybridising to either
PT   the PERV-A or PERV-B env gene - useful in the detection of
PT   retroviruses, and their subtypes, in a sample of porcine/human
PT   tissue.
XX
PS   Claim 3; Page 21-23; 36pp; English.
XX
CC   The present sequence encodes a Pig endogenous retrovirus (PERV)-B
CC   envelope protein. PERV exists in two different subtypes, PERV-A and
CC   PERV-B. The differences are reflected in sequence divergence in the
CC   envelope genes. Probes and primers can be derived from the envelope
CC   (env) genes of PERV-A and PERV-B. The probes and primers are used in
CC   a method to detect retroviruses in a sample of porcine/human tissue,
CC   particularly primary porcine tissue and human cell lines that have been
CC   cultivated in the presence of a porcine cell line, or human tissue from
CC   a patient with a xenotransplant. Subtype of PERV in a sample containing
CC   one of the PERV env genes can also be determined.
XX
SQ   Sequence 3482 BP; 927 A; 854 C; 867 G; 834 T; 0 other;


  Query Match            100.0%;  Score 18;  DB 20;  Length 3482;
  Best Local Similarity  100.0%;  Pred. No. 2.6;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          ||||||||||||||||||
Db   3439 CCACAGTCGTACACCACG 3422


RESULT   4
AAC67022/c
ID   AAC67022 standard; DNA; 4918 BP.
XX
AC   AAC67022;
XX
DT   27-MAR-2001  (first entry)
XX
```

DE PERV env protein coding sequence SEQ ID NO: 22.
XX
KW Xenotransplantation; infectious agent; vaccine; ds.
XX
OS Porcine endogenous retrovirus.
XX
PN WO200071726-A1.
XX
PD 30-NOV-2000.
XX
PF 24-MAY-2000; 2000WO-US14296.
XX
PR 24-MAY-1999; 99US-0135631.
XX
PA (MAYO-) MAYO MEDICAL VENTURES.
XX
PI Federspiel MJ;
XX
DR WPI; 2001-032041/04.
XX
PT Inhibiting or preventing infectious agent transmission in mammalian
PT transplant recipients, by introducing recombinant DNA comprising DNA
PT encoding extracellular proteins of the agent into donor cells, such as
PT swine cells -
XX
PS Claim 16; Page 109-111; 144pp; English.
XX
CC The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX
SQ Sequence 4918 BP; 1382 A; 1173 C; 1183 G; 1180 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 4918;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccacagtcgtacaccacg 18
||||||||||||||||||
Db 4781 CCACAGTCGTACACCACG 4764

RESULT 5
AAC67021/c
ID AAC67021 standard; DNA; 6076 BP.
XX
AC AAC67021;
XX
DT 27-MAR-2001 (first entry)
XX
DE PERV env protein coding sequence SEQ ID NO: 21.
XX
KW Xenotransplantation; infectious agent; vaccine; ds.
XX
OS Porcine endogenous retrovirus.
XX
PN WO200071726-A1.
XX
PD 30-NOV-2000.
XX
PF 24-MAY-2000; 2000WO-US14296.
XX
PR 24-MAY-1999; 99US-0135631.
XX
PA (MAYO-) MAYO MEDICAL VENTURES.
XX
PI Federspiel MJ;
XX
DR WPI; 2001-032041/04.
XX
PT Inhibiting or preventing infectious agent transmission in mammalian
PT transplant recipients, by introducing recombinant DNA comprising DNA
PT encoding extracellular proteins of the agent into donor cells, such as
PT swine cells -
XX
PS Claim 16; Page 107-109; 144pp; English.
XX
CC The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX
SQ Sequence 6076 BP; 1613 A; 1512 C; 1609 G; 1342 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 6076;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccacagtcgtacaccacg 18
||||||||||||||||||
Db 617 CCACAGTCGTACACCACG 600

RESULT 6
AAC67032/c
ID AAC67032 standard; DNA; 6076 BP.
XX
AC AAC67032;
XX
DT 27-MAR-2001 (first entry)
XX
DE PERV env protein coding sequence SEQ ID NO: 32.
XX
KW Xenotransplantation; infectious agent; vaccine; ds.
XX
OS Porcine endogenous retrovirus.
XX
PN WO200071726-A1.
XX
PD 30-NOV-2000.
XX
PF 24-MAY-2000; 2000WO-US14296.
XX
PR 24-MAY-1999; 99US-0135631.
XX
PA (MAYO-) MAYO MEDICAL VENTURES.
XX
PI Federspiel MJ;
XX
DR WPI; 2001-032041/04.
XX
PT Inhibiting or preventing infectious agent transmission in mammalian
PT transplant recipients, by introducing recombinant DNA comprising DNA
PT encoding extracellular proteins of the agent into donor cells, such as
PT swine cells -
XX
PS Claim 16; Page 117-119; 144pp; English.
XX
CC The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX
SQ Sequence 6076 BP; 1619 A; 1505 C; 1598 G; 1354 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 6076;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ccacagtcgtacaccacg 18
          ||||||||||||||||||
Db    617 CCACAGTCGTACACCACG 600


RESULT   7
AAF77726/c
ID   AAF77726 standard; DNA; 7333 BP.
XX
AC   AAF77726;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Defective retroviral genome isolated from PK-15 cell line.
XX
KW   Retrovirus; graft transplantation; xenotransplantation; PK-15 cell line;
KW   ss.
XX
OS   Unidentified.
XX
PN   US6190861-B1.
XX
PD   20-FEB-2001.
XX
PF   13-DEC-1996;   96US-0766528.
XX
PR   14-DEC-1995;   95US-0572645.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Fishman JA;
XX
DR   WPI; 2001-256211/26.
DR   P-PSDB; AAB73282, AAB73283, AAB73284.
XX
PT   Assessing risk of endogenous retroviruses in clinical practice and in
PT   xenotransplantation, comprises using probe sequences derived from swine
PT   or miniature swine retroviral genome -
XX
PS   Claim 1; Fig 2; 127pp; English.
XX
CC   The present invention relates to a method for screening a cell or tissue
CC   for the presence or expression of a retrovirus (RV), comprising
CC   contacting a target nucleic acid from the cell or tissue with a second
CC   nucleic acid from the present invention (e.g. the present sequence or a
CC   fragment thereof). The method is useful for RV detection and to assess
CC   graft transplantation risk. Screening of animals allows the elimination
CC   of donors with active replication of known viruses. Inactive proviruses
CC   can be detected and inactivated, allowing identification and elimination
CC   of potential human pathogens derived from swine in a manner not possible
CC   in the outbred human organ donor population and is important to the
CC   development of human xenotransplantation.
XX
SQ   Sequence 7333 BP; 1984 A; 1788 C; 1885 G; 1676 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 22;  Length 7333;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          ||||||||||||||||||
Db   7279 CCACAGTCGTACACCACG 7262


RESULT   8
AAC67019/c
ID   AAC67019 standard; DNA; 7362 BP.
XX
AC   AAC67019;
XX
DT   27-MAR-2001  (first entry)
XX
DE   PERV env protein coding sequence SEQ ID NO: 19.
XX
KW   Xenotransplantation; infectious agent; vaccine; ds.
XX
OS   Porcine endogenous retrovirus.
XX
PN   WO200071726-A1.
XX
PD   30-NOV-2000.
XX
PF   24-MAY-2000; 2000WO-US14296.
XX
PR   24-MAY-1999;   99US-0135631.
XX
PA   (MAYO-) MAYO MEDICAL VENTURES.
XX
PI   Federspiel MJ;
XX
DR   WPI; 2001-032041/04.
XX
PT   Inhibiting or preventing infectious agent transmission in mammalian
PT   transplant recipients, by introducing recombinant DNA comprising DNA
PT   encoding extracellular proteins of the agent into donor cells, such as
PT   swine cells -
XX
PS   Claim 16; Page 101-104; 144pp; English.
XX
CC   The present invention provides a method to prevent the transmission of
CC   infectious agents during xenotransplantation. This involves introducing
CC   to donor swine cells a recombinant DNA encoding a peptide fragment from
CC   the infectious agent, and then introducing these cells into the
CC   transplant recipient.
XX
SQ   Sequence 7362 BP; 1997 A; 1821 C; 1881 G; 1663 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 22;  Length 7362;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          ||||||||||||||||||
Db    682 CCACAGTCGTACACCACG 665


RESULT   9
AAT74883/c
ID   AAT74883 standard; cDNA; 7393 BP.
XX
AC   AAT74883;
XX
DT   09-FEB-1998  (first entry)
XX
DE   Porcine retrovirus cDNA (defective).
XX
KW   Retrovirus; porcine; GAG protein; POL protein; ENV protein;
KW   xenotransplantation; infectious; provirus; organ transplant; donor;
KW   activated virus; PCR; ss.
XX
OS   Porcine retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             598..2172
FT                   /*tag=  a
FT   mat_peptide     598..2169
FT                   /*tag=  b
FT                   /note= "putative GAG protein"
FT   CDS             2320..4737
FT                   /*tag=  c
FT                   /note= "putative POL coding region (partial) as
FT                         described in the specification"
FT   mat_peptide     2320..3522
FT                   /*tag=  d
```

```
FT                   /note= "putative POL protein (partial)"
FT   mat_peptide     3516..4328
FT                   /*tag=  e
FT                   /note= "putative POL protein (partial)"
FT   mat_peptide     4332..4748
FT                   /*tag=  f
FT   CDS             4738..6725
FT                   /*tag=  g
FT                   /note= "putative ENV coding region (partial) as
FT                          described in the specification"
FT   mat_peptide     4752..6722
FT                   /*tag=  h
FT                   /note= "ENV protein (partial)"
XX
PN   WO9721836-A1.
XX
PD   19-JUN-1997.
XX
PF   13-DEC-1996;   96WO-US19680.
XX
PR   14-DEC-1995;   95US-0572645.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Fishman JA;
XX
DR   WPI; 1997-332804/30.
DR   P-PSDB; AAW32091-W32095.
XX
PT   New nucleic acid from porcine retro:viruses - used for detecting
PT   viruses in transplant or other tissue and for assessing risk of
PT   transmitting infection to graft recipient
XX
PS   Claim 16; Fig 2; 128pp; English.
XX
CC   This cDNA sequence represents a defective purified swine retrovirus
CC   found in PK-15 cells containing the putative coding regions for viral
CC   GAG, POL and ENV proteins. There are a few in frame stop codons and
CC   apparent frame shifts in the given coding sequence which alter features
CC   of the translation. This sequence and PCR fragments generated from the
CC   sequence (see AAT74812-T74882) could be used to screen organs for the
CC   presence of porcine retroviruses prior to xenotransplantation.
CC   Transplantation can increase the likelihood of retroviral activation if
CC   intact and infectious proviruses are present. The porcine retroviral
CC   sequence can be used to generate probes to determine the level (e.g.
CC   copy number) of intact (i.e. potentially replicating) porcine provirus
CC   sequences in a strain of xenograft transplantation donors. It can be
CC   used to detect mutations, genetic lesions or viral recombinants and to
CC   determine the histological localisation of activated retrovirus. Using
CC   Polymerase Chain Reaction DNA Quantitation (PDQ) on blood mononuclear
CC   cells, infectivity titration and susceptibility testing can be
CC   performed. Ultimately animal donors without intact porcine retroviral
CC   sequences or a lower copy number of viral elements could be selected.
XX
SQ   Sequence 7393 BP; 2004 A; 1801 C; 1896 G; 1692 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 18;  Length 7393;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          ||||||||||||||||||
Db   7339 CCACAGTCGTACACCACG 7322


RESULT  10
AAC67023/c
ID   AAC67023 standard; DNA; 7873 BP.
XX
AC   AAC67023;
XX
DT   27-MAR-2001  (first entry)
XX
DE   PERV env protein coding sequence SEQ ID NO: 23.
XX
KW   Xenotransplantation; infectious agent; vaccine; ds.
XX
OS   Porcine endogenous retrovirus.
XX
PN   WO200071726-A1.
XX
PD   30-NOV-2000.
XX
PF   24-MAY-2000; 2000WO-US14296.
XX
PR   24-MAY-1999;   99US-0135631.
XX
PA   (MAYO-) MAYO MEDICAL VENTURES.
XX
PI   Federspiel MJ;
XX
DR   WPI; 2001-032041/04.
XX
PT   Inhibiting or preventing infectious agent transmission in mammalian
PT   transplant recipients, by introducing recombinant DNA comprising DNA
PT   encoding extracellular proteins of the agent into donor cells, such as
PT   swine cells -
XX
PS   Claim 16; Page 112-115; 144pp; English.
XX
CC   The present invention provides a method to prevent the transmission of
CC   infectious agents during xenotransplantation. This involves introducing
CC   to donor swine cells a recombinant DNA encoding a peptide fragment from
CC   the infectious agent, and then introducing these cells into the
CC   transplant recipient.
XX
SQ   Sequence 7873 BP; 2200 A; 1913 C; 1962 G; 1798 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 22;  Length 7873;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          ||||||||||||||||||
Db   7708 CCACAGTCGTACACCACG 7691


RESULT  11
AAT74884/c
ID   AAT74884 standard; cDNA; 7892 BP.
XX
AC   AAT74884;
XX
DT   09-FEB-1998  (first entry)
XX
DE   Miniature swine retrovirus cDNA.
XX
KW   Retrovirus; porcine; GAG protein; POL protein; ENV protein;
KW   xenotransplantation; infectious; provirus; organ transplant; donor;
KW   activated virus; PCR; ss.
XX
OS   Porcine retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             585..2159
FT                   /*tag=  a
FT   mat_peptide     585..2156
FT                   /*tag=  b
FT                   /note= "putative GAG protein"
FT   CDS             2307..5744
FT                   /*tag=  c
FT   mat_peptide     2307..5741
FT                   /*tag=  d
FT                   /note= "putative POL protein"
```

```
FT   CDS             5620..7536
FT                   /*tag=  e
FT   mat_peptide     5620..7533
FT                   /*tag=  f
FT                   /note= "putative ENV protein"
XX
PN   WO9721836-A1.
XX
PD   19-JUN-1997.
XX
PF   13-DEC-1996;   96WO-US19680.
XX
PR   14-DEC-1995;   95US-0572645.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Fishman JA;
XX
DR   WPI; 1997-332804/30.
DR   P-PSDB; AAW32096-W32098.
XX
PT   New nucleic acid from porcine retro:viruses - used for detecting
PT   viruses in transplant or other tissue and for assessing risk of
PT   transmitting infection to graft recipient
XX
PS   Claim 22; Fig 3; 128pp; English.
XX
CC   This cDNA sequence represents a porcine retrovirus from miniature swine
CC   containing the putative coding regions for viral GAG, POL and ENV
CC   proteins. This sequence and PCR fragments generated from it
CC   (see AAT74812-T74882) can be used to screen organs for the presence of
CC   porcine retroviruses prior to xenotransplantation. Transplantation can
CC   increase the likelihood of retroviral activation if intact and
CC   infectious proviruses are present. The porcine retroviral sequence can be
CC   used to generate probes to determine the level (e.g. copy number) of
CC   intact (i.e. potentially replicating) porcine provirus sequences in a
CC   strain of xenograft transplantation donors. It can be used to detect
CC   mutations, genetic lesions or viral recombinants and also to determine
CC   the histological localisation of activated retrovirus. Using Polymerase
CC   Chain Reaction DNA Quantitation (PDQ) on blood mononuclear cells,
CC   infectivity titration and susceptibility testing can be performed.
CC   Ultimately animal donors without intact porcine retroviral sequences or a
CC   lower copy number of viral elements could be selected.
XX
SQ   Sequence 7892 BP; 2191 A; 1915 C; 1980 G; 1806 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 18;  Length 7892;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccacg 18
           ||||||||||||||||||
Db      19 CCACAGTCGTACACCACG 2


RESULT  12
AAT74811/c
ID   AAT74811 standard; cDNA; 8060 BP.
XX
AC   AAT74811;
XX
DT   11-FEB-1998  (first entry)
XX
DE   Porcine retrovirus Tsukuba-1 cDNA.
XX
KW   Retrovirus; porcine; GAG protein; POL protein; ENV protein;
KW   xenotransplantation; infectious; provirus; organ transplant; donor;
KW   activated virus; Tsukuba-1; PCR; ss.
XX
OS   Porcine retrovirus.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             86..2002
FT                   /*tag=  a
FT                   /note= "putative ENV protein"
FT   CDS             3112..4686
FT                   /*tag=  b
FT                   /note= "putative GAG protein"
FT   CDS             4871..8060
FT                   /*tag=  c
FT                   /note= "putative POL protein (partial)"
XX
PN   WO9721836-A1.
XX
PD   19-JUN-1997.
XX
PF   13-DEC-1996;   96WO-US19680.
XX
PR   14-DEC-1995;   95US-0572645.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Fishman JA;
XX
DR   WPI; 1997-332804/30.
XX
PT   New nucleic acid from porcine retro:viruses - used for detecting
PT   viruses in transplant or other tissue and for assessing risk of
PT   transmitting infection to graft recipient
XX
PS   Claim 1; Fig 1; 128pp; English.
XX
CC   This sequence represents the purified porcine retroviral cDNA
CC   sequence of Tsukuba-1 and contains the putative coding regions for viral
CC   proteins GAG, POL and ENV. This sequence and PCR fragments generated
CC   from the sequence (see AAT74812-T74882) could be used to screen organs
CC   for porcine retroviruses prior to xenotransplantation. Transplantation
CC   can increase the likelihood of retroviral activation if intact and
CC   infectious proviruses are present. The porcine retroviral sequence can be
CC   used to generate probes to determine the level (e.g. copy number) of
CC   intact (i.e. potentially replicating) porcine provirus sequences in a
CC   strain of xenograft transplantation donors. It can be used to detect
CC   mutations, genetic lesions or viral recombinants and to determine the
CC   histological localisation of activated retrovirus. Using Polymerase Chain
CC   Reaction DNA Quantitation (PDQ) on blood mononuclear cells, infectivity
CC   titration and susceptibility testing can be performed. Ultimately animal
CC   donors without intact porcine retroviral sequences or with a lower copy
CC   number of viral elements could be selected.
XX
SQ   Sequence 8060 BP; 2233 A; 1959 C; 2012 G; 1856 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 18;  Length 8060;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccacg 18
           ||||||||||||||||||
Db    2546 CCACAGTCGTACACCACG 2529


RESULT  13
AAF77725/c
ID   AAF77725 standard; cDNA; 8060 BP.
XX
AC   AAF77725;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Tsukuba-1 cDNA.
XX
KW   Tsukuba-1; retrovirus; graft transplantation; xenotransplantation; ss.
XX
OS   Porcine retrovirus.
XX
```

```
PN   US6190861-B1.
XX
PD   20-FEB-2001.
XX
PF   13-DEC-1996;   96US-0766528.
XX
PR   14-DEC-1995;   95US-0572645.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Fishman JA;
XX
DR   WPI; 2001-256211/26.
XX
PT   Assessing risk of endogenous retroviruses in clinical practice and in
PT   xenotransplantation, comprises using probe sequences derived from swine
PT   or miniature swine retroviral genome -
XX
PS   Claim 1; Fig 1; 127pp; English.
XX
CC   The present invention relates to a method for screening a cell or tissue
CC   for the presence or expression of a retrovirus (RV), comprising
CC   contacting a target nucleic acid from the cell or tissue with a second
CC   nucleic acid from the present invention (e.g. the present sequence or a
CC   fragment thereof). The method is useful for RV detection and to assess
CC   graft transplantation risk. Screening of animals allows the elimination
CC   of donors with active replication of known viruses. Inactive proviruses
CC   can be detected and inactivated, allowing identification and elimination
CC   of potential human pathogens derived from swine in a manner not possible
CC   in the outbred human organ donor population and is important to the
CC   development of human xenotransplantation.
XX
SQ   Sequence 8060 BP; 2233 A; 1959 C; 2012 G; 1856 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 22;  Length 8060;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccacg 18
           ||||||||||||||||||
Db    2546 CCACAGTCGTACACCACG 2529


RESULT  14
AAF77727/c
ID   AAF77727 standard; DNA; 8132 BP.
XX
AC   AAF77727;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of a retovirus found in miniature swine.
XX
KW   Retrovirus; graft transplantation; xenotransplantation; miniature swine;
KW   ss.
XX
OS   Unidentified.
XX
PN   US6190861-B1.
XX
PD   20-FEB-2001.
XX
PF   13-DEC-1996;   96US-0766528.
XX
PR   14-DEC-1995;   95US-0572645.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Fishman JA;
XX
DR   WPI; 2001-256211/26.
DR   P-PSDB; AAB73285, AAB73286, AAB73287.
XX
PT   Assessing risk of endogenous retroviruses in clinical practice and in
PT   xenotransplantation, comprises using probe sequences derived from swine
PT   or miniature swine retroviral genome -
XX
PS   Claim 1; Fig 3; 127pp; English.
XX
CC   The present invention relates to a method for screening a cell or tissue
CC   for the presence or expression of a retrovirus (RV), comprising
CC   contacting a target nucleic acid from the cell or tissue with a second
CC   nucleic acid from the present invention (e.g. the present sequence or a
CC   fragment thereof). The method is useful for RV detection and to assess
CC   graft transplantation risk. Screening of animals allows the elimination
CC   of donors with active replication of known viruses. Inactive proviruses
CC   can be detected and inactivated, allowing identification and elimination
CC   of potential human pathogens derived from swine in a manner not possible
CC   in the outbred human organ donor population and is important to the
CC   development of human xenotransplantation.
XX
SQ   Sequence 8132 BP; 2248 A; 1977 C; 2037 G; 1870 T; 0 other;


  Query Match             100.0%;  Score 18;  DB 22;  Length 8132;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccacg 18
           ||||||||||||||||||
Db    8080 CCACAGTCGTACACCACG 8063


RESULT  15
AAV09699/c
ID   AAV09699 standard; DNA; 8196 BP.
XX
AC   AAV09699;
XX
DT   19-MAY-1998  (first entry)
XX
DE   Porcine retrovirus DNA encoding, GAG, POL and ENV.
XX
KW   Porcine retrovirus; PoEV; POL protein; ENV protein; GAG protein;
KW   vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
XX
OS   Porcine retrovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             576..2126
FT                   /*tag=  a
FT                   /product= GAG protein
FT                   /note= "virion core polypeptide"
FT   CDS             2143..5733
FT                   /*tag=  b
FT                   /product= POL protein
FT                   /note= "polymerase peptide sequence as given in
FT                         specification"
FT   CDS             5606..7576
FT                   /*tag=  c
FT                   /product= ENV protein
FT                   /note= "envelope protein"
XX
PN   WO9740167-A1.
XX
PD   30-OCT-1997.
XX
PF   18-APR-1997;   97WO-GB01087.
XX
PR   10-FEB-1997;   97GB-0002668.
PR   19-APR-1996;   96GB-0008164.
XX
PA   (IMUT-) IMUTRAN LTD.
PA   (QONE-) Q-ONE BIOTECH LTD.
XX
```

PI Galbraith DN, Haworth C, Lees GM, Smith KT;
XX
DR WPI; 1997-535851/49.
XX
PT Polynucleotide encoding porcine retrovirus expression product -
PT useful to develop products for use in vaccines, diagnosis and
PT xeno-transplantation
XX
PS Claim 4; Fig 2; 69pp; English.
XX
CC This DNA sequence encodes the porcine retrovirus (PoEV) virion core
CC polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins. These
CC proteins can be used to develop viral vaccines, antisense nucleic acids,
CC ribozymes and other antiviral agents. They can also be used in
CC xeno-transplantation technology and as diagnostic tools.
XX
SQ Sequence 8196 BP; 2165 A; 2061 C; 2147 G; 1820 T; 3 other;

Query Match 100.0%; Score 18; DB 18; Length 8196;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccacagtcgtacaccacg 18
||||||||||||||||||
Db 8130 CCACAGTCGTACACCACG 8113

Search completed: February 24, 2002, 03:05:32
Job time: 19498 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 02:56:31 ; Search time 197.25 Seconds
(without alignments)
3811.949 Million cell updates/sec

Title: US-09-171-553B-1
Perfect score: 3320
Sequence: 1 gaattcgcggccgcgtcgac.....aaaaaaaaaaaaaaaaaaaa 3320

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1091	32.9	7333	4	US-08-766-528-2	Sequence 2, Appli
	2	786	23.7	2462	3	US-09-111-085-1	Sequence 1, Appli
	3	704	21.2	8132	4	US-08-766-528-3	Sequence 3, Appli
	4	375	11.3	8060	4	US-08-766-528-1	Sequence 1, Appli
	5	321	9.7	3482	3	US-09-111-085-3	Sequence 3, Appli
	6	265	8.0	445	4	US-09-376-781-7	Sequence 7, Appli
	7	35	1.1	8655	3	US-09-075-272-1	Sequence 1, Appli
	8	30	0.9	836	4	US-09-352-990-7	Sequence 7, Appli
	9	30	0.9	1279	4	US-09-277-716-31	Sequence 31, Appl
	10	29	0.9	735	1	US-08-798-000-3	Sequence 3, Appli
	11	29	0.9	1198	3	US-09-248-335-27	Sequence 27, Appl
	12	29	0.9	1234	1	US-08-798-000-1	Sequence 1, Appli
	13	29	0.9	2338	4	US-09-232-200-66	Sequence 66, Appl
	14	29	0.9	2338	4	US-09-232-197-66	Sequence 66, Appl
	15	29	0.9	7308	4	US-09-011-745-3	Sequence 3, Appli
	16	29	0.9	7308	4	US-09-011-745-4	Sequence 4, Appli
	17	29	0.9	7616	4	US-09-011-745-2	Sequence 2, Appli
c	18	29	0.9	8202	1	US-08-258-420-13	Sequence 13, Appl
	19	29	0.9	8332	3	US-08-850-961-1	Sequence 1, Appli
	20	28	0.8	140	1	US-08-628-417-5	Sequence 5, Appli
	21	28	0.8	216	1	US-08-686-878A-34	Sequence 34, Appl
	22	28	0.8	240	1	US-08-628-417-6	Sequence 6, Appli
	23	28	0.8	330	4	US-09-078-294-24	Sequence 24, Appl
c	24	28	0.8	595	4	US-09-385-982-278	Sequence 278, App
	25	28	0.8	698	4	US-09-376-781-18	Sequence 18, Appl
	26	28	0.8	719	1	US-08-375-346A-1	Sequence 1, Appli
	27	28	0.8	719	2	US-08-467-123B-1	Sequence 1, Appli
	28	28	0.8	949	5	PCT-US91-08254-1	Sequence 1, Appli
c	29	28	0.8	949	5	PCT-US91-08254-2	Sequence 2, Appli
	30	28	0.8	1132	3	US-08-894-731-3	Sequence 3, Appli
	31	28	0.8	1578	4	US-09-416-050A-1	Sequence 1, Appli
	32	28	0.8	1578	4	US-09-664-800-1	Sequence 1, Appli
	33	28	0.8	1578	4	US-09-665-309-1	Sequence 1, Appli
	34	28	0.8	1578	4	US-09-661-569-1	Sequence 1, Appli
	35	28	0.8	1964	3	US-08-468-856B-7	Sequence 7, Appli
	36	28	0.8	1964	3	US-08-468-859A-7	Sequence 7, Appli
	37	28	0.8	2000	4	US-09-376-781-25	Sequence 25, Appl
	38	28	0.8	2000	4	US-09-376-781-30	Sequence 30, Appl
	39	28	0.8	3291	4	US-09-318-448-12	Sequence 12, Appl
	40	27	0.8	43	2	US-08-686-599A-14	Sequence 14, Appl
	41	27	0.8	104	2	US-08-803-899-8	Sequence 8, Appli
	42	27	0.8	153	4	US-09-244-794A-3	Sequence 3, Appli
	43	27	0.8	153	4	US-09-247-190-3	Sequence 3, Appli
	44	27	0.8	159	4	US-09-244-794A-17	Sequence 17, Appl
	45	27	0.8	159	4	US-09-247-190-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-766-528-2
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-2

Query Match 32.9%; Score 1091; DB 4; Length 7333;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1364 aaaggggttgcttacctcagcagggagggaaataaagaacaaagaggaaattctaagcc 1423

```
Db    3517 AAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAGGAAATTCTAAGCC 3576

Qy    1424 tattagaagccttacatttgccaaaaaggctagctattatacactgtcctggacatcaga 1483
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3577 TATTAGAAGCCTTACATTTGCCAAAAAGGCTAGCTATTATACACTGTCCTGGACATCAGA 3636

Qy    1484 aagccaaagatctcatatctagagggaaccagatggctgaccgggttgccaagcaggcag 1543
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3637 AAGCCAAAGATCTCATATCTAGAGGGAACCAGATGGCTGACCGGGTTGCCAAGCAGGCAG 3696

Qy    1544 cccaggctgttaaccttctgcctataatagaaacgcccaaagccccagaacccagacgac 1603
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3697 CCCAGGCTGTTAACCTTCTGCCTATAATAGAAACGCCCAAAGCCCCAGAACCCAGACGAC 3756

Qy    1604 agtacaccctagaagactggcaagagataaaaaagatagaccagttctctgagactccgg 1663
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3757 AGTACACCCTAGAAGACTGGCAAGAGATAAAAAAGATAGACCAGTTCTCTGAGACTCCGG 3816

Qy    1664 aggggacctgctatacctcatatgggaaggaaatcctgccccacaaagaagggttagaat 1723
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3817 AGGGGACCTGCTATACCTCATATGGGAAGGAAATCCTGCCCCACAAAGAAGGGTTAGAAT 3876

Qy    1724 atgtccaacagatacatcgtctaacccacctaggaactaaacacctgcagcagttggtca 1783
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3877 ATGTCCAACAGATACATCGTCTAACCCACCTAGGAACTAAACACCTGCAGCAGTTGGTCA 3936

Qy    1784 gaacatccccttatcatgttctgaggctaccaggagtggctgactcggtggtcaaacatt 1843
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3937 GAACATCCCCTTATCATGTTCTGAGGCTACCAGGAGTGGCTGACTCGGTGGTCAAACATT 3996

Qy    1844 gtgtgccctgccagctggttaatgctaatccttccagaatacctccaggaaagagactaa 1903
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3997 GTGTGCCCTGCCAGCTGGTTAATGCTAATCCTTCCAGAATACCTCCAGGAAAGAGACTAA 4056

Qy    1904 ggggaagccacccaggcgctcactgggaagtggacttcactgaggtaaagccggctaaat 1963
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4057 GGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAAT 4116

Qy    1964 acggaaacaaatatctattggtttttgtagacacctttttcaggatgggtagaggcttatc 2023
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4117 ACGGAAACAAATATCTATTGGTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATC 4176

Qy    2024 ctactaagaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttccaa 2083
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4177 CTACTAAAAAAGAGACTTCAACCGTGGTGGCTAAGAAAATACTGGAGGAAATTTTTCCAA 4236

Qy    2084 gatttggaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccaggtaa 2143
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4237 GATTTGGAATACCTAAGGTAATAGGGTCAGACAATGGTCCAGCTTTCGTTGCCCAGGTAA 4296

Qy    2144 gtcagggactggccaagatattggggattgattggaaactgcattgtgcatacagacccc 2203
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    4297 GTCAGGGACTGGCCAAGATATTGGGGATTGATTGAAAACTGCATTGTGCATACAGACCCC 4356

Qy    2204 aaagctcaggacaggtagagaggatgaatagaaccattaaagagacccttaccaaattga 2263
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4357 AAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACCCTTACCAAATTGA 4416

Qy    2264 ccacagagactggcattaatgattggatggctctcctgccctttgtgctttttagggtga 2323
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4417 CCACAGAGACTGGCATTAATGATTGGATGGCTCTCCTGCCCTTTGTGCTTTTTAGGGTGA 4476

Qy    2324 ggaacacccctggacagtttgggctgacccccctatgaattgctctacgggggaccccccc 2383
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    4477 GGAACACCCCTGGACAGTTTGGGCTGACCCCCTATAAATTGCTCTACGGGGGACCCCCCC 4536

Qy    2384 cgttggcagaaattgcctttgcacatagtgctgatgtgctgctttcccagcctttgttct 2443
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4537 CGTTGGCAGAAATTGCCTTTGCACATAGTGCTGATGTGCTGCTTTCCCAGCCTTTGTTCT 4596

Qy    2444 ctaggctcaaggcgctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcct 2503
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4597 CTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAGCGAGCGTGGAAGCAGCTCCGGGAGGCCT 4656
```

```
Qy    2504 actcaggaggagacttgcaagttccacatcgcttccaagttggagattcagtctatgtta 2563
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4657 ACTCAGGAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTTGGAGATTCAGTCTATGTTA 4716

Qy    2564 gacgccaccgtgcaggaaacctcgagactcggtggaagggaccttatctctgtacttttga 2623
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    4717 GACGCCACCGTGCAGGAAACCTCGAGACTCGGTAGAAGGGACCTTATCTCGTACTTTTGA 4776

Qy    2624 ccacaccaacggctgtgaaagtcgaaggaatcccc 2658
           ||||||||||||||||||||||||||||||||||
Db    4777 CCACACCAACGGCTGTGAAAGTCGAAGGAATCCCC 4811
```

RESULT 2
US-09-111-085-1
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2462
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-1

Query Match 23.7%; Score 786; DB 3; Length 2462;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 836; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    2459 tcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagact 2518
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 tcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagact 60

Qy    2519 tgcaagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgcag 2578
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 tgcaagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgcag 120

Qy    2579 gaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacggctg 2638
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacggctg 180

Qy    2639 tgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccggcgccacctc 2698
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 tgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccggcgccacctc 240

Qy    2699 ccgattcggggtggaaagccgaaaagactgaaaatccccttaagcttcgcctccatcgcg 2758
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ccgattcggggtggaaagccgaaaagactgaaaatccccttaagcttcgcctccatcgcg 300

Qy    2759 tggttccttactctgtcaataactcctcaagttaatggtaaacgccttgtggacagcccg 2818
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 tggttccttactctgtcaataactcctcaagttaatggtaaacgccttgtggacagcccg 360

Qy    2819 aactcccataaaacccttatctctcacctggttacttactgactccggtacaggtattaat 2878
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 aactcccataaaacccttatctctcacctggttacttactgactccggtacaggtattaat 420

Qy    2879 attaacagcactcaaggggaggctcccttggggacctggtggcctgaattatatgtctgc 2938
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 attaacagcactcaaggggaggctcccttggggacctggtggcctgaattatatgtctgc 480
```

```
Qy   2939 cttcgatcagtaatccctggtctcaatgaccaggccacacccccgatgtactccgtgct 2998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 cttcgatcagtaatccctggtctcaatgaccaggccacacccccgatgtactccgtgct 540

Qy   2999 tacgggttttacgtttgcccaggacccccaaataatgaagaatattgtggaaatcctcag 3058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 tacgggttttacgtttgcccaggacccccaaataatgaagaatattgtggaaatcctcag 600

Qy   3059 gatttcctttgcaagcaatggagctgcataacttctaatgatgggaattggaaatggcca 3118
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 gatttcttttgcaagcaatggagctgcataacttctaatgatgggaattggaaatggcca 660

Qy   3119 gtctctcagcaagacagagtaagttactcttttgttaacaatcctaccagttataatcaa 3178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gtctctcagcaagacagagtaagttactcttttgttaacaatcctaccagttataatcaa 720

Qy   3179 tttaattatggccatgggagatggaaagattggcaacagcgggtacaaaaagatgtacga 3238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 tttaattatggccatgggagatggaaagattggcaacagcgggtacaaaaagatgtacga 780

Qy   3239 aataagcaaataagctgtcattcgttagacctagattacttaaaaataagtttcact 3295
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 aataagcaaataagctgtcattcgttagacctagattacttaaaaataagtttcact 837
```

```
RESULT   3
US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8132 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-3


  Query Match             21.2%;  Score 704;  DB 4;  Length 8132;
  Best Local Similarity   98.7%;  Pred. No. 2.8e-299;
  Matches 2204;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;
```

```
Qy     22 gatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaatgg 81
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2970 GATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAACCACTTTTTGCCTTCGAATGG 3029

Qy     82 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3030 AGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTC 3089

Qy    142 aagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatc 201
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   3090 AAGAACTCCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATC 3149

Qy    202 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3150 CAACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACC 3209

Qy    262 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 321
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3210 AAACAGGACTGCTTAGAAGGCACGAAGGCACTACTGCTGGAATTGTCTGACCTAGGCTAC 3269

Qy    322 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3270 AGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGGTAACATACTTGGGGTACAGT 3329

Qy    382 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 441
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3330 TTGCGGGACGGGCAGCGATGGCTGACGGAGGCACGGAAGAAAACTGTAGTCCAGATACCG 3389

Qy    442 gccccaaccacagccaaacaagtgagagagtttttggggacagctggatttttgcagactg 501
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3390 GCCCCAACCACAGCCAAACAAATGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTG 3449

Qy    502 tggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 561
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3450 TGGATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGG 3509

Qy    562 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 621
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3510 GAATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATGCTATCAAAAAGGCCCTGCTG 3569

Qy    622 agcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgag 681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3570 AGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCTTTACCCTTTATGTGGATGAG 3629

Qy    682 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 741
          |||||||||||||||| |||||||||||||||||||||||||||||||||||| ||||||||
Db   3630 CGTAAGGGAGTAGCCCGGGGAGTTTTAACCCAAACCCTAGGACCATGGAGAAGACCTGTC 3689

Qy    742 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 801
          |||||||||||||||||| |||||||||||||||||||||||| |||| ||||||||
Db   3690 GCCTACCTGTCAAAGAAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATGCCTGAAGGCT 3749

Qy    802 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatata 861
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3750 ATCGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAGAATATA 3809

Qy    862 actgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatggatg 921
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3810 ACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAGACCGATGGATG 3869

Qy    922 accaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgct 981
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   3870 ACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCACAGAGAGGGTCACGTTCGCT 3929

Qy    982 ccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgact 1041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3930 CCACCAGCCGCTCTCAACCCTGCCACTCTTCTGCCTGAAGAGACTGATGAACCAGTGACT 3989

Qy   1042 catgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacata 1101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3990 CATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATA 4049

Qy   1102 ccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggt 1161
```

```
Db   4050 CCGCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGGT 4109

Qy   1162 aagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctg 1221
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   4110 AAGAGGATGGCTGGGGCGGCGGTGGTGGACGGGACCCGCACGATCTGGGCCAGCAGCCTG 4169

Qy   1222 ccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctg 1281
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   4170 CCGGAAGGAACTTCAGCACAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCTG 4229

Qy   1282 gccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcacac 1341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4230 GCCGAAGGGAAATCCATAAACATTTATACGGACAGCAGGTATGCCTTTGCGACTGCACAC 4289

Qy   1342 gtacacggggccatctataaacaaaggggggttgcttacctcagcagggagggaaataaag 1401
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4290 GTACATGGGGCCATCTATAAACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAAG 4349

Qy   1402 aacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctatt 1461
          |||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db   4350 AACAAAGAGGAAATTCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCTATT 4409

Qy   1462 atacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggct 1521
          |||||||||||||||||||||||||||||| |||||||||||||| ||||| ||||||||||
Db   4410 ATACACTGTCCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCAGATGGCT 4469

Qy   1522 gaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgccc 1581
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||||
Db   4470 GACCGGGTTGCCAAGCAGGCAGCCCAGGGTGTTAACCTTCTGCCTATAATAGAAATGCCC 4529

Qy   1582 aaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagata 1641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4530 AAAGCCCCAGAACCCAGACGACAGTACACCCTAGAAGACTGGCAAGAGATAAAAAAGATA 4589

Qy   1642 gaccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaatcctg 1701
          |||||||||||||||||||| |||||||||||||||||| ||||||||||||||||||||
Db   4590 GACCAGTTCTCTGAGACTCCGGAAGGGACCTGCTATACCTCAGATGGGAAGGAAATCCTG 4649

Qy   1702 ccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaact 1761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4650 CCCCACAAAGAAGGGTTAGAATATGTCCAACAGATACATCGTCTAACCCACCTAGGAACT 4709

Qy   1762 aaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtg 1821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4710 AAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGGCTACCAGGAGTG 4769

Qy   1822 gctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccaga 1881
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4770 GCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGGTTAATGCTAATCCTTCCAGA 4829

Qy   1882 atacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttc 1941
          || |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4830 ATGCCTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTC 4889

Qy   1942 actgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacaccttt 2001
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   4890 ACTGAGGTAAAGCCGGCTAAATACGGAAACAAATACCTATTGGTTTTTGTAGACACCTTT 4949

Qy   2002 tcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaa 2061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   4950 TCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAAAAA 5009

Qy   2062 atactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggt 2121
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5010 ATACTGGAAGAAATTTTTCCAAGATTTGGAATACCTAAGGTAATAGGGTCAGACAATGGT 5069

Qy   2122 ccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaa 2181
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5070 CCAGCTTTTGTTGCCCAGGTAAGTCAGGGACTGGCCAAGATATTGGGGATTGATTGGAAA 5129

Qy   2182 ctgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccatt 2241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   5130 CTGCATTGTGCATACAGACCCCAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATT 5189

Qy   2242 aaagagacccttac 2255
          ||||||||||||||
Db   5190 AAAGAGACCCTTAC 5203
```

RESULT 4
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-1

Query Match 11.3%; Score 375; DB 4; Length 8060;
Best Local Similarity 98.7%; Pred. No. 3.5e-155;
Matches 1395; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

```
Qy     22 gatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaatgg 81
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5495 GATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAACCACTTTTTGCCTTCGAATGG 5554

Qy     82 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5555 AGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTC 5614

Qy    142 aagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatc 201
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   5615 AAGAACTCCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATC 5674

Qy    202 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5675 CAACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACC 5734

Qy    262 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 321
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
```

```
Db   5735 AAACAGGACTGCTTAGAAGGCACGAAGGCACTACTGCTGGAATTGTCTGACCTAGGCTAC 5794

Qy    322 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5795 AGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGGTAACATACTTGGGGTACAGT 5854

Qy    382 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 441
          || |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5855 TTACGGGACGGGCAGCGATGGCTGACGGAGGCACGGAAGAAAACTGTAGTCCAGATACCG 5914

Qy    442 gccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagactg 501
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   5915 GCCCCAACCACAGCCAAACAAATGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTG 5974

Qy    502 tggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5975 TGGATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGG 6034

Qy    562 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 621
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6035 GAATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATGCTATCAAAAAGGCCCTGCTG 6094

Qy    622 agcgcacctgctctggccctccctgacgtaactaaaccctttacccctttatgtggatgag 681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6095 AGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCTTTACCCTTTATGTGGATGAG 6154

Qy    682 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 741
          |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db   6155 CGTAAGGGAGTAGCCCGGGGAGTTTTAACCCAAACCCTAGGACCATGGAGAAGACCTGTC 6214

Qy    742 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 801
          |||||||||||||||||| |||||||||||||||||||||||||| |||| |||||||||
Db   6215 GCCTACCTGTCAAAGAAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATGCCTGAAGGCT 6274

Qy    802 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggac-agaatat 860
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   6275 ATCGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAAGAATAT 6334

Qy    861 aactgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatggat 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6335 AACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAGACCGATGGAT 6394

Qy    921 gaccaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgc 980
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   6395 GACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCACAGAGAGGGTCACGTTCGC 6454

Qy    981 tccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgac 1040
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6455 TCCACCAACCGCTCTCAACCCTGCCACTCTTCTGCCTGAAGAGACTGATGAACCAGTGAC 6514

Qy   1041 tcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacat 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6515 TCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACAT 6574

Qy   1101 accgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaagg 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6575 ACCGCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGG 6634

Qy   1161 taagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcct 1220
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   6635 TAAGAGGATGGCTGGGGCGGCGGTGGTGGACGGGACCCGCACGATCTGGGCCAGCAGCCT 6694

Qy   1221 gccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggct 1280
          ||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   6695 GCCGGGAGGAACTTCAGCACAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCT 6754

Qy   1281 ggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcaca 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6755 GGCCGAAGGGAAATCCATAAACATTTATACGGACAGCAGGTATGCCTTTGCGACTGCACA 6814

Qy   1341 cgtacacggggccatctataaacaaaggggggttgcttacctcagcagggagggaaataaa 1400
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6815 CGTACATGGGGCCATCTATAAACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAA 6874

Qy   1401 gaacaaagaggaaattctaagcctattagaagcc 1434
          ||||||||||||||||||||||||||||||||
Db   6875 GAACAAAGAGGAAATTCTAAGCCTATTAGAAGCC 6908


RESULT   5
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
;  APPLICANT: Stoye, Jonathan P
;  APPLICANT: Weiss, Robin A
;  TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
;  TITLE OF INVENTION: specific sequences
;  FILE REFERENCE: 4238/75168
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 3482
;   TYPE: DNA
;   ORGANISM: Porcine retrovirus
US-09-111-085-3


  Query Match             9.7%;  Score 321;  DB 3;  Length 3482;
  Best Local Similarity   99.2%;  Pred. No. 1.7e-131;
  Matches  521;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy   1767 cctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctga 1826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 cctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctga 65

Qy   1827 ctcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacc 1886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 ctcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacc 125

Qy   1887 tccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactga 1946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    126 tccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactga 185

Qy   1947 ggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcagg 2006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 ggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcagg 245

Qy   2007 atgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatact 2066
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    246 atgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaaaaaaatact 305

Qy   2067 ggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagc 2126
          |||||||||||||| |||||||||||||||||||| ||||||||||||||||||||||
Db    306 ggaggaaatttttccgagatttggaatacctaaggtaatcgggtcagacaatggtccagc 365

Qy   2127 tttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgca 2186
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 ttttgttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgca 425

Qy   2187 ttgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaaga 2246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 ttgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaaga 485

Qy   2247 gacccttaccaaattgaccacagagactggcattaatgattggat 2291
          |||||||||||||||||||||||||||||||||||||||||||
Db    486 gacccttaccaaattgaccacagagactggcattaatgattggat 530


RESULT   6
US-09-376-781-7
; Sequence 7, Application US/09376781
```

```
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 445
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: sequence of a portion of the PERV-A env gene
;   OTHER INFORMATION: corresponding to nucleotides 364-809 of GenBank
;   OTHER INFORMATION: No. 6261806 Y12238.
US-09-376-781-7


  Query Match             8.0%;  Score 265;  DB 4;  Length 445;
  Best Local Similarity   99.5%;  Pred. No. 6.5e-107;
  Matches  435;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy   2823 cccataaacccttatctctcacctggttacttactgactccggtacaggtattaatatta 2882
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cccataaacccttatctctcacctggttacttactgactccggtacaggtattaatatta 60

Qy   2883 acagcactcaaggggaggctcccttggggacctggtggcctgaattatatgtctgccttc 2942
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 acagcactcaaggggaggctcccttggggacctggtggcctgaattatatgtctgccttc 120

Qy   2943 gatcagtaatccctggtctcaatgaccaggccacacccccgatgtactccgtgcttacg 3002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gatcagtaatccctggtctcaatgaccaggccacacccccgatgtactccgtgcttacg 180

Qy   3003 ggttttacgtttgcccaggacccccaaataatgaagaatattgtggaaatcctcaggatt 3062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ggttttacgtttgcccaggacccccaaataatgaagaatattgtggaaatcctcaggatt 240

Qy   3063 tcctttgcaagcaatggagctgcataacttctaatgatg-ggaattggaaatggccagtc 3121
           || |||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    241 tcttttgcaagcaatggagctgcataacttctaatgatgaggaattggaaatggccagtc 300

Qy   3122 tctcagcaagacagagttaagttactcttttgttaacaatcctaccagttataatcaattt 3181
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 tctcagcaagacagagttaagttactcttttgttaacaatcctaccagttataatcaattt 360

Qy   3182 aattatggccatgggagatggaaagattggcaacagcgggtacaaaaagatgtacgaaat 3241
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 aattatggccatgggagatggaaagattggcaacagcgggtacaaaaagatgtacgaaat 420

Qy   3242 aagcaaataagctgtca 3258
           |||||||||||||||||
Db    421 aagcaaataagctgtca 437


RESULT   7
US-09-075-272-1
; Sequence 1, Application US/09075272
; Patent No. 6136598
;  GENERAL INFORMATION:
;    APPLICANT:  MILLER, A. DUSTY
;    APPLICANT:  WOLGAMOT, GREG
;    APPLICANT:  BONHAM, LYNN
;    TITLE OF INVENTION:  MUS DUNNI ENDOGENOUS RETROVIRAL
;    TITLE OF INVENTION:  PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,272
;      FILING DATE:  08-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/046,140
;      FILING DATE:  09-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poor, Brian W.
;      REGISTRATION NUMBER:  32,928
;      REFERENCE/DOCKET NUMBER:  14538A-003710
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-075-272-1


  Query Match             1.1%;  Score 35;  DB 3;  Length 8655;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-06;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2203 caaagctcaggacaggtagagaggatgaatagaac 2237
           |||||||||||||||||||||||||||||||||||
Db   5249 CAAAGCTCAGGACAGGTAGAGAGGATGAATAGAAC 5283


RESULT   8
US-09-352-990-7
; Sequence 7, Application US/09352990
; Patent No. 6255090
; GENERAL INFORMATION:
;  APPLICANT: Famodu, Layo O.
;  APPLICANT: Orozco, Buddy
;  APPLICANT: Rafalski, Antoni
;  TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
;  FILE REFERENCE: BB-1191
;  CURRENT APPLICATION NUMBER: US/09/352,990
;  CURRENT FILING DATE: 1999-07-14
;  EARLIER APPLICATION NUMBER: 60/092,866
;  EARLIER FILING DATE: July 15, 1998
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
;   LENGTH: 836
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-352-990-7


  Query Match             0.9%;  Score 30;  DB 4;  Length 836;
  Best Local Similarity   100.0%;  Pred. No. 0.00053;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy   3291 tcactaaaaaaaaaaaaaaaaaaaaaaaaa 3320
          |||||||||||||||||||||||||||||
Db    742 tcactaaaaaaaaaaaaaaaaaaaaaaaaa 771


RESULT   9
US-09-277-716-31
; Sequence 31, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
;  APPLICANT: Bryan, Bruce
;  APPLICANT: Szent-Gyorgyi, Christopher
;  APPLICANT: PROLUME, LTD.
;  TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
;  CURRENT APPLICATION NUMBER: US/09/277,716A
;  CURRENT FILING DATE: 1999-03-26
;  EARLIER APPLICATION NUMBER: 60/102,939
;  EARLIER FILING DATE: 1998-10-01
;  EARLIER APPLICATION NUMBER: 60/089,367
;  EARLIER FILING DATE: 1998-06-15
;  EARLIER APPLICATION NUMBER: 60/079,624
;  EARLIER FILING DATE: 1998-03-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 1279
;   TYPE: DNA
;   ORGANISM: Ptilosarcus gurneyi
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7)..(720)
;   FEATURE:
;   OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert B)
US-09-277-716-31


  Query Match             0.9%;  Score 30;  DB 4;  Length 1279;
  Best Local Similarity   100.0%;  Pred. No. 0.00051;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3291 tcactaaaaaaaaaaaaaaaaaaaaaaaaa 3320
          |||||||||||||||||||||||||||||
Db   1237 tcactaaaaaaaaaaaaaaaaaaaaaaaaa 1266


RESULT  10
US-08-798-000-3
; Sequence 3, Application US/08798000
; Patent No. 5766945
;  GENERAL INFORMATION:
;    APPLICANT:  Miller, A. Dusty
;    TITLE OF INVENTION:  10A1 RETROVIRAL PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/798,000
;      FILING DATE:  12-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/011,564
;      FILING DATE:  13-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  14538A-002010
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-467-9600
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  735 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-798-000-3


  Query Match             0.9%;  Score 29;  DB 1;  Length 735;
  Best Local Similarity   100.0%;  Pred. No. 0.0015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2185 cattgtgcatacagaccccaaagctcagg 2213
          |||||||||||||||||||||||||||||
Db     96 CATTGTGCATACAGACCCCAAAGCTCAGG 124


RESULT  11
US-09-248-335-27
; Sequence 27, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
;  APPLICANT: MCGONIGLE, BRIAN
;  APPLICANT: O'KEEF, DANIEL
;  TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE: CL-1128-A
;  CURRENT APPLICATION NUMBER: US/09/248,335
;  CURRENT FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: 08/924,759
;  EARLIER FILING DATE: 1997-September-05
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 27
;   LENGTH: 1198
;   TYPE: DNA
;   ORGANISM: maize
US-09-248-335-27


  Query Match             0.9%;  Score 29;  DB 3;  Length 1198;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3292 cactaaaaaaaaaaaaaaaaaaaaaaaaa 3320
          |||||||||||||||||||||||||||||
Db   1109 cactaaaaaaaaaaaaaaaaaaaaaaaaa 1137


RESULT  12
US-08-798-000-1
; Sequence 1, Application US/08798000
; Patent No. 5766945
;  GENERAL INFORMATION:
;    APPLICANT:  Miller, A. Dusty
;    TITLE OF INVENTION:  10A1 RETROVIRAL PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/798,000
;      FILING DATE:  12-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/011,564
;      FILING DATE:  13-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  14538A-002010
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-467-9600
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1234 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-798-000-1


  Query Match             0.9%;  Score 29;  DB 1;  Length 1234;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2185 cattgtgcatacagaccccaaagctcagg 2213
           ||||||||||||||||||||||||||||||
Db      96 CATTGTGCATACAGACCCCAAAGCTCAGG 124


RESULT  13
US-09-232-200-66
; Sequence 66, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 66
;   LENGTH: 2338
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-200-66


  Query Match             0.9%;  Score 29;  DB 4;  Length 2338;
  Best Local Similarity   100.0%;  Pred. No. 0.0013;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3292 cactaaaaaaaaaaaaaaaaaaaaaaaaa 3320
           |||||||||||||||||||||||||||||
Db    2295 cactaaaaaaaaaaaaaaaaaaaaaaaaa 2323


RESULT  14
US-09-232-197-66
; Sequence 66, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 66
;   LENGTH: 2338
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-197-66


  Query Match             0.9%;  Score 29;  DB 4;  Length 2338;
  Best Local Similarity   100.0%;  Pred. No. 0.0013;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3292 cactaaaaaaaaaaaaaaaaaaaaaaaaa 3320
           |||||||||||||||||||||||||||||
Db    2295 cactaaaaaaaaaaaaaaaaaaaaaaaaa 2323


RESULT  15
US-09-011-745-3
; Sequence 3, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 7308
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
US-09-011-745-3


  Query Match             0.9%;  Score 29;  DB 4;  Length 7308;
  Best Local Similarity   100.0%;  Pred. No. 0.0012;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    2185 cattgtgcatacagaccccaaagctcagg 2213
           ||||||||||||||||||||||||||||||
Db    6178 cattgtgcatacagaccccaaagctcagg 6206
```

Search completed: February 24, 2002, 07:23:41
Job time: 16030 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:23:41 ; Search time 197.25 Seconds
(without alignments)
9425.389 Million cell updates/sec

Title: US-09-171-553B-3
Perfect score: 8209
Sequence: 1 gtggtgtacgactgtgggcc.....aaaaaaaaaaaaaaaaaaa 8209

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

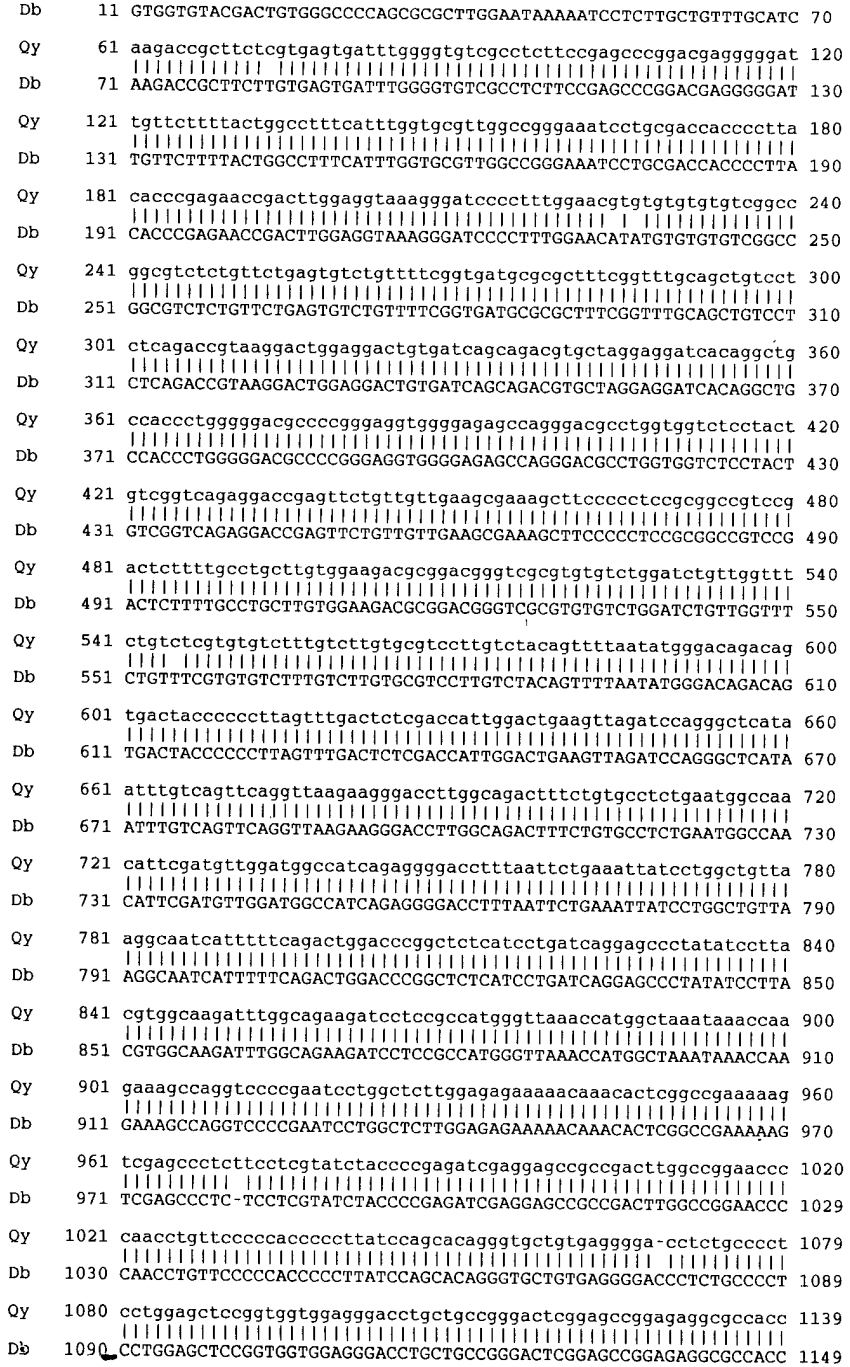
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2967	36.1	7333	4	US-08-766-528-2	Sequence 2, Appli
	2	1466	17.9	3482	3	US-09-111-085-3	Sequence 3, Appli
	3	704	8.6	8132	4	US-08-766-528-3	Sequence 3, Appli
	4	457	5.6	2462	3	US-09-111-085-1	Sequence 1, Appli
	5	375	4.6	8060	4	US-08-766-528-1	Sequence 1, Appli
	6	278	3.4	278	4	US-09-376-781-10	Sequence 10, Appl
	7	86	1.0	276	4	US-09-376-781-8	Sequence 8, Appli
	8	35	0.4	8655	3	US-09-075-272-1	Sequence 1, Appli
	9	29	0.4	735	1	US-08-798-000-3	Sequence 3, Appli
	10	29	0.4	1234	1	US-08-798-000-1	Sequence 1, Appli
	11	29	0.4	7308	4	US-09-011-745-3	Sequence 3, Appli
	12	29	0.4	7308	4	US-09-011-745-4	Sequence 4, Appli
	13	29	0.4	7616	4	US-09-011-745-2	Sequence 2, Appli
c	14	29	0.4	8202	1	US-08-258-420-13	Sequence 13, Appl
	15	29	0.4	8332	3	US-08-850-961-1	Sequence 1, Appli
	16	28	0.3	698	4	US-09-376-781-18	Sequence 18, Appl
	17	28	0.3	949	5	PCT-US91-08254-1	Sequence 1, Appli
c	18	28	0.3	949	5	PCT-US91-08254-2	Sequence 2, Appli
	19	28	0.3	1493	4	US-09-376-781-24	Sequence 24, Appl
	20	28	0.3	1736	4	US-09-360-197-13	Sequence 13, Appl
	21	28	0.3	2000	4	US-09-376-781-25	Sequence 25, Appl
	22	28	0.3	2000	4	US-09-376-781-30	Sequence 30, Appl
	23	27	0.3	1587	3	US-09-108-020-11	Sequence 11, Appl
c	24	26	0.3	843	2	US-08-831-570-2	Sequence 2, Appli
	25	26	0.3	1098	3	US-09-248-335-35	Sequence 35, Appl
c	26	26	0.3	1275	3	US-08-725-532A-2	Sequence 2, Appli
	27	26	0.3	4032	1	US-08-126-587C-8	Sequence 8, Appli
	28	26	0.3	80595	4	US-09-078-294-3	Sequence 3, Appli
c	29	25	0.3	25	4	US-09-376-781-23	Sequence 23, Appl
	30	25	0.3	945	4	US-09-268-364-3	Sequence 3, Appli
	31	25	0.3	1000	4	US-09-357-251-19	Sequence 19, Appl
	32	25	0.3	1001	1	US-08-728-259A-10	Sequence 10, Appl
	33	25	0.3	1001	2	US-08-473-486-10	Sequence 10, Appl
	34	25	0.3	1026	2	US-08-713-000-6	Sequence 6, Appli
	35	25	0.3	1026	2	US-08-975-316-6	Sequence 6, Appli
	36	25	0.3	1026	4	US-09-211-710-6	Sequence 6, Appli
	37	25	0.3	1055	1	US-08-702-344-18	Sequence 18, Appl
	38	25	0.3	1066	2	US-08-605-106-3	Sequence 3, Appli
	39	25	0.3	1622	4	US-09-334-601-9	Sequence 9, Appli
	40	25	0.3	1737	4	US-09-416-050A-7	Sequence 7, Appli
	41	25	0.3	1737	4	US-09-664-800-7	Sequence 7, Appli
	42	25	0.3	1737	4	US-09-665-309-7	Sequence 7, Appli
	43	25	0.3	1737	4	US-09-661-569-7	Sequence 7, Appli
	44	25	0.3	1826	2	US-08-808-931-15	Sequence 15, Appl
	45	25	0.3	1826	3	US-08-808-323-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-766-528-2
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-2

Query Match 36.1%; Score 2967; DB 4; Length 7333;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3507; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

Qy 1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

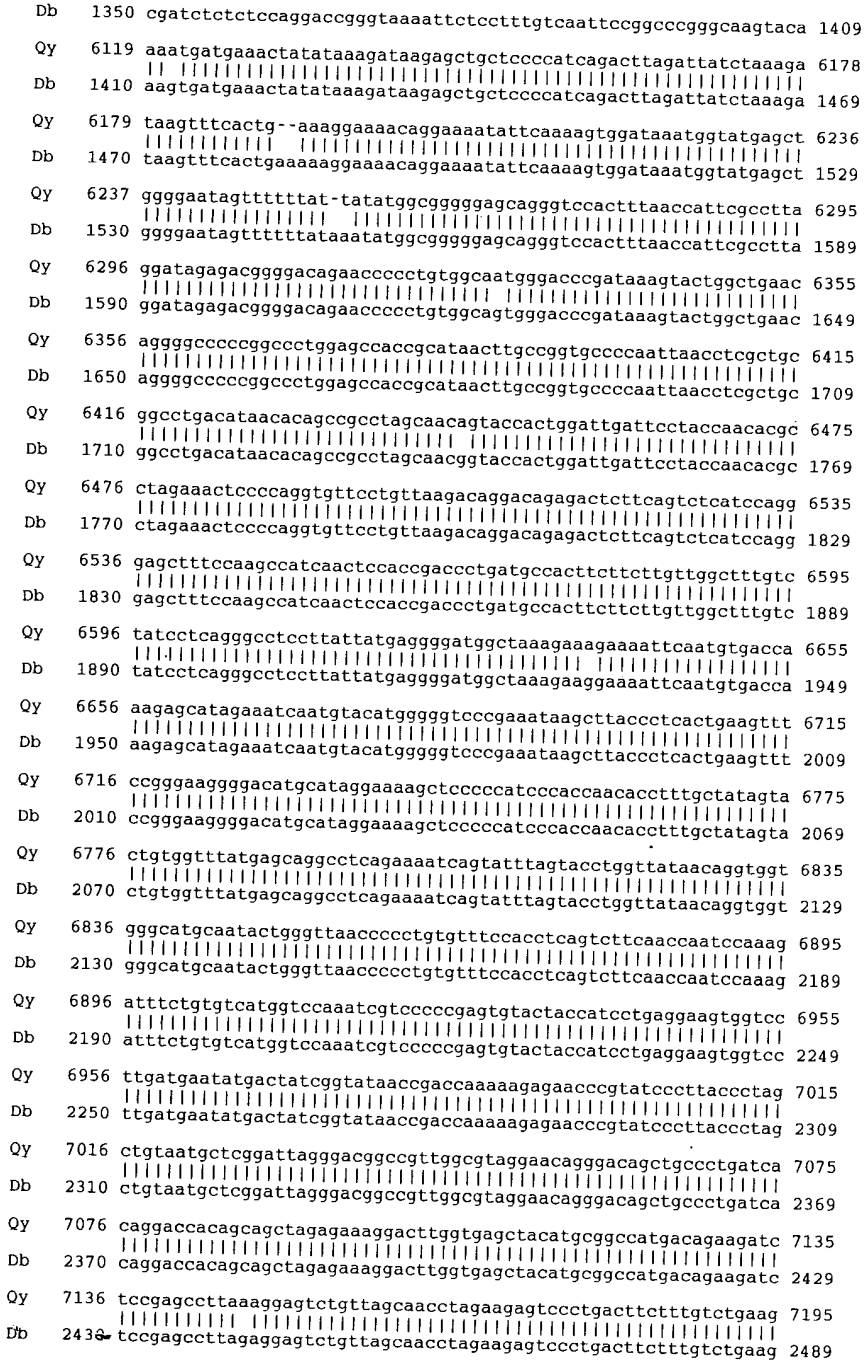

Db 11 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 70
Qy 61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggat 120
Db 71 AAGACCGCTTCTTGTGAGTGATTTGGGGTGTCGCCTCTTCCGAGCCCGGACGAGGGGGAT 130
Qy 121 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctta 180
Db 131 TGTTCTTTTACTGGCCTTTCATTTGGTGCGTTGGCCGGGAAATCCTGCGACCACCCCTTA 190
Qy 181 cacccgagaaccgacttggaggtaaagggatccccttggaacgtgtgtgtgtgtcggcc 240
Db 191 CACCCGAGAACCGACTTGGAGGTAAAGGGATCCCCTTTGGAACATATGTGTGTGTCGGCC 250
Qy 241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
Db 251 GGCGTCTCTGTTCTGAGTGTCTGTTTTCGGTGATGCGCGCTTTCGGTTTGCAGCTGTCCT 310
Qy 301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
Db 311 CTCAGACCGTAAGGACTGGAGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAGGCTG 370
Qy 361 ccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 420
Db 371 CCACCCTGGGGGACGCCCCGGGAGGTGGGGAGAGCCAGGGACGCCTGGTGGTCTCCTACT 430
Qy 421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 480
Db 431 GTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCG 490
Qy 481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
Db 491 ACTCTTTTGCCTGCTTGTGGAAGACGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGTTT 550
Qy 541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
Db 551 CTGTTTCGTGTGTCTTTGTCTTGTGCGTCCTTGTCTACAGTTTTAATATGGGACAGACAG 610
Qy 601 tgactacccccctttagtttgactctcgaccattggactgaagttagatccagggctcata 660
Db 611 TGACTACCCCCCTTAGTTTGACTCTCGACCATTGGACTGAAGTTAGATCCAGGGCTCATA 670
Qy 661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
Db 671 ATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGCCTCTGAATGGCCAA 730
Qy 721 cattcgatgttggatggccatcagaggggacctttaattctgaaattatcctggctgtta 780
Db 731 CATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCTGAAATTATCCTGGCTGTTA 790
Qy 781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatccttta 840
Db 791 AGGCAATCATTTTTCAGACTGGACCCGGCTCTCATCCTGATCAGGAGCCCTATATCCTTA 850
Qy 841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
Db 851 CGTGGCAAGATTTGGCAGAAGATCCTCCGCCATGGGTTAAACCATGGCTAAATAAACCAA 910
Qy 901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
Db 911 GAAAGCCAGGTCCCCGAATCCTGGCTCTTGGAGAGAAAAACAAACACTCGGCCGAAAAAG 970
Qy 961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
Db 971 TCGAGCCCTC-TCCTCGTATCTACCCCGAGATCGAGGAGCCGCCGACTTGGCCGGAACCC 1029
Qy 1021 caacctgttcccccaccccctttatccagcacagggtgctgtgagggga-cctctgcccct 1079
Db 1030 CAACCTGTTCCCCCACCCCCTTATCCAGCACAGGGTGCTGTGAGGGGACCCTCTGCCCCT 1089
Qy 1080 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1139
Db 1090 CCTGGAGCTCCGGTGGTGGAGGGACCTGCTGCCGGGACTCGGAGCCGGAGAGGCGCCACC 1149
Qy 1140 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1199
Db 1150 CCGGAGCGGACAGACGAGATCGCGATATTACCGCTGCGCACCTATGGCCCTCCCATGCCA 1209
Qy 1200 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1259
Db 1210 GGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGG 1269
Qy 1260 aaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtcc 1319
Db 1270 AAAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGAGTCC 1329
Qy 1320 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1379
Db 1330 CTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTC 1389
Qy 1380 acaaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgac 1439
Db 1390 ACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAAAAAAAATGTTCCTGGGGCCGAC 1449
Qy 1440 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1499
Db 1450 GGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGATTTCCCTTGACTCGCCCCGGT 1509
Qy 1500 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1559
Db 1510 TGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCTGGTG 1569
Qy 1560 gcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatg 1619
Db 1570 GCGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAGGTGATG 1629
Qy 1620 cagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcgg 1679
Db 1630 CAGGGACCGAACGAACCTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGG 1689
Qy 1680 ttcaccccttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcatt 1739
Db 1690 TTCACCCCTTTTGATCCTACCTCAGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTCATT 1749
Qy 1740 gggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggct 1799
Db 1750 GGGCAGTCGGCTCTGGATATCAGGAAGAAACTTCAGAGACTGGAAGGGTTACAGGAGGCT 1809
Qy 1800 gagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagag 1859
Db 1810 GAGTTACGTGATCTAGTGAGAGAGGCAGAGAAGGTGTATTACAGAAGGGAGACAGAAGAG 1869
Qy 1860 gagaaggaacagagaaaagaaaaggagagagaagaaaagggaggaaagacgtgatagacgg 1919
Db 1870 GAGAAGGAACAGAGAAAAGAAAAGGAGAGAGAAGAAAGGGAGGAAAGACGTGATAGACGG 1929
Qy 1920 caagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggag 1979
Db 1930 CAAGAGAAGAATTTGACTAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAG 1989
Qy 1980 agagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaat 2039
Db 1990 AGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAAT 2049
Qy 2040 aggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagg 2099
Db 2050 AGGACCCCACTCGACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGGACACTGGGCAAGG 2109
Qy 2100 aactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagat 2159
Db 2110 AACTGCCCCAAGAAGGGAAACAAAGGACCGAAGGTCCTAGCTCTAGAAGAAGATAAAGAT 2169
Qy 2160 taggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtggagggg 2219
Db 2170 TAGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGAGGGG 2229

```
Qy   2220 caaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccatta 2279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2230 CAACCAGTTGAGTTCCTGGTTGATACCGGAGCGGAGCATTCAGTGCTGCTACAACCATTA 2289

Qy   2280 ggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatcca 2339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2290 GGAAAACTAAAAGAAAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACGGCAGTATCCA 2349

Qy   2340 tggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtc 2399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2350 TGGACTACCCGAAGAACCGTTGACTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTC 2409

Qy   2400 atccctgagtgcccagtacccctctaggtagagacttactgaccaagatgggagctcaa 2459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2410 ATCCCTGAGTGCCCAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAA 2469

Qy   2460 atttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttg 2519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2470 ATTTCTTTTGAACAAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCACTGTGTTG 2529

Qy   2520 accctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagat 2579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2530 ACCCTCCAATTAGATGATGAATATCGACTATATTCTCCCCAAGTAAAGCCTGATCAAGAT 2589

Qy   2580 atacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttg 2639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2590 ATACAGTCCTGGTTGGAGCAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGGTTTG 2649

Qy   2640 gcaaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtc 2699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2650 GCAAAGCAAGTTCCCCCACAGGTTATTCAACTGAAGGCCAGTGCTACACCAGTATCAGTC 2709

Qy   2700 agacagtacccccttgagtagagaggctcgagaaggaatttggccgcatgttcaaagatta 2759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2710 AGACAGTACCCCTTGAGTAGAGAGGCTCGAGAAGGAATTTGGCCGCATGTTCAAAGATTA 2769

Qy   2760 atccaacagggcatcctagttcctgtccaatccccttggaatactcccctgctaccggtt 2819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2770 ATCCAACAGGGCATCCTAGTTCCTGTCCAATCCCCTTGGAATACTCCCCTGCTACCGGTT 2829

Qy   2820 aggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaagg 2879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2830 AGGAAGCCTGGGACCAATGATTATCGACCAGTACAGGACTTGAGAGAGGTCAATAAAAGG 2889

Qy   2880 gtgcaggacatacacccaacggtcccgaacccttataacctcttgagcgccctcccgcct 2939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2890 GTGCAGGACATACACCCAACGGTCCCGAACCCTTATAACCTCTTGAGCGCCCTCCCGCCT 2949

Qy   2940 gaacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacac 2999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2950 GAACGGAACTGGTACACAGTATTGGACTTAAAAGATGCCTTCTTCTGCCTGAGATTACAC 3009

Qy   3000 cccactagccaaccacttttttgccttcgaatggagagatccaggtacgggaagaaccggg 3059
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3010 CCCACTAGCCAACCACTTTTTTACCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGG 3069

Qy   3060 cagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaa 3119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3070 CAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGACGAA 3129

Qy   3120 gccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccag 3179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3130 GCCCTACACAGGGACCTGGCCAACTTCAGGATCCAACACCCTCAGGTGACCCTCCTCCAG 3189

Qy   3180 tacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaag 3239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3190 TACGTGGATGACCTGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGTACGAAG 3249

Qy   3240 gcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagatt 3299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3250 GCACTACTGCTGGAATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATT 3309

Qy   3300 tgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacg 3359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3310 TGCAGGAGAGAGGTAACATACTTGGGGTACAGTTTGCGGGGCGGGCAGCGATGGCTGACG 3369

Qy   3360 gaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgaga 3419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3370 GAGGCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAGTGAGA 3429

Qy   3420 gagtttttggggacagctggattttgcagactgtggatcccggggtttgcgaccttagca 3479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3430 GAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTTAGCA 3489

Qy   3480 gccccactctacccgctaaccaaagaaaaaggggg 3514
          |||||||||||||||||||||||||||||||||
Db   3490 GCCCCACTCTACCCGCTAACCAAAGAAAAAGGGGG 3524
```

RESULT 2
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

Query Match 17.9%; Score 1466; DB 3; Length 3482;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2536; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

```
Qy   5639 ggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcg 5698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    930 ggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcg 989

Qy   5699 cctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgcctta 5758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    990 cctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgcctta 1049

Qy   5759 tagacagctcgaaccccccatagacctttatcccttacctggctgattattgaccctgata 5818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1050 tagacagctcgaaccccccatagacctttatcccttacctggctgattattgaccctgata 1109

Qy   5819 cgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac 5878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1110 cgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac 1169

Qy   5879 tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtcc 5938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtcc 1229

Qy   5939 gtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctg 5998
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 gtagttatgggttctattgctgcccaggcacagagaaagagaaatactgtgggggttctg 1289

Qy   5999 gggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggc 6058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 gggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatggc 1349

Qy   6059 cgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagtaca 6118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



RESULT 3
US-08-766-528-3
; Sequence 3, Application US/08766528

```
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8132 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-3


  Query Match             8.6%;  Score 704;  DB 4;  Length 8132;
  Best Local Similarity   98.7%;  Pred. No. 2.5e-293;
  Matches 2204;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy   2973 gatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaatgg 3032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2970 GATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAACCACTTTTTGCCTTCGAATGG 3029

Qy   3033 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 3092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3030 AGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTC 3089

Qy   3093 aagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatc 3152
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   3090 AAGAACTCCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATC 3149

Qy   3153 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 3212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3150 CAACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACC 3209

Qy   3213 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 3272
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3210 AAACAGGACTGCTTAGAAGGCACGAAGGCACTACTGCTGGAATTGTCTGACCTAGGCTAC 3269

Qy   3273 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 3332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3270 AGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGGTAACATACTTGGGGTACAGT 3329

Qy   3333 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 3392
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3330 TTGCGGGACGGGCAGCGATGGCTGACGGAGGCACGGAAGAAAACTGTAGTCCAGATACCG 3389

Qy   3393 gccccaaccacagccaaacaagtgagagagtttttggggacagctggatttttgcagactg 3452
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3390 GCCCCAACCACAGCCAAACAAATGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTG 3449

Qy   3453 tggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 3512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3450 TGGATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGG 3509

Qy   3513 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 3572
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3510 GAATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATGCTATCAAAAAGGCCCTGCTG 3569

Qy   3573 agcgcacctgctctggccctccctgacgtaactaaaacccttaccctttatgtggatgag 3632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3570 AGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCTTTACCCTTTATGTGGATGAG 3629

Qy   3633 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 3692
          |||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db   3630 CGTAAGGGAGTAGCCCGGGGAGTTTTAACCCAAACCCTAGGACCATGGAGAAGACCTGTC 3689

Qy   3693 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 3752
          |||||||||||||||||| |||||||||||||||||||||||||| |||| |||||||||
Db   3690 GCCTACCTGTCAAAGAAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATGCCTGAAGGCT 3749

Qy   3753 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatata 3812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3750 ATCGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAGAATATA 3809

Qy   3813 actgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatggatg 3872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3810 ACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAGACCGATGGATG 3869

Qy   3873 accaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgct 3932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   3870 ACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCACAGAGAGGGTCACGTTCGCT 3929

Qy   3933 ccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgact 3992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3930 CCACCAGCCGCTCTCAACCCTGCCACTCTTCTGCCTGAAGAGACTGATGAACCAGTGACT 3989

Qy   3993 catgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacata 4052
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3990 CATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATA 4049

Qy   4053 ccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggt 4112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4050 CCGCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGGT 4109

Qy   4113 aagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctg 4172
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   4110 AAGAGGATGGCTGGGGCGGCGGTGGTGGACGGGACCCGCACGATCTGGGCCAGCAGCCTG 4169

Qy   4173 ccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctg 4232
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   4170 CCGGAAGGAACTTCAGCACAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCTG 4229

Qy   4233 gccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcacac 4292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4230 GCCGAAGGGAAATCCATAAACATTTATACGGACAGCAGGTATGCCTTTGCGACTGCACAC 4289

Qy   4293 gtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaag 4352
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4290 GTACATGGGGCCATCTATAAACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAAG 4349

Qy   4353 aacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctatt 4412
          |||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db   4350 AACAAAGAGGAAATTCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCTATT 4409

Qy   4413 atacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggct 4472
          |||||||||||||||||||||||||| |||||||||||||| ||||| ||||||||||||
Db   4410 ATACACTGTCCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCAGATGGCT 4469

Qy   4473 gaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgccc 4532
```

```
Db   4470 GACCGGGTTGCCAAGCAGGCAGCCCAGGGTGTTAACCTTCTGCCTATAATAGAAATGCCC 4529

Qy   4533 aaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaaagata 4592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4530 AAAGCCCCAGAACCCAGACGACAGTACACCCTAGAAGACTGGCAAGAGATAAAAAAGATA 4589

Qy   4593 gaccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaatcctg 4652
          |||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||
Db   4590 GACCAGTTCTCTGAGACTCCGGAAGGGACCTGCTATACCTCAGATGGGAAGGAAATCCTG 4649

Qy   4653 ccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaact 4712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4650 CCCCACAAAGAAGGGTTAGAATATGTCCAACAGATACATCGTCTAACCCACCTAGGAACT 4709

Qy   4713 aaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtg 4772
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4710 AAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGGCTACCAGGAGTG 4769

Qy   4773 gctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccaga 4832
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4770 GCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGGTTAATGCTAATCCTTCCAGA 4829

Qy   4833 atacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttc 4892
          || |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4830 ATGCCTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTC 4889

Qy   4893 actgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacaccttt 4952
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   4890 ACTGAGGTAAAGCCGGCTAAATACGGAAACAAATACCTATTGGTTTTTGTAGACACCTTT 4949

Qy   4953 tcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaa 5012
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   4950 TCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAAAAA 5009

Qy   5013 atactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggt 5072
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5010 ATACTGGAAGAAATTTTTCCAAGATTTGGAATACCTAAGGTAATAGGGTCAGACAATGGT 5069

Qy   5073 ccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaa 5132
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5070 CCAGCTTTTGTTGCCCAGGTAAGTCAGGGACTGGCCAAGATATTGGGGATTGATTGGAAA 5129

Qy   5133 ctgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccatt 5192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5130 CTGCATTGTGCATACAGACCCCAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATT 5189

Qy   5193 aaagagacccttac 5206
          ||||||||||||||
Db   5190 AAAGAGACCCTTAC 5203
```

RESULT 4
US-09-111-085-1
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2462
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-1

Query Match 5.6%; Score 457; DB 3; Length 2462;
Best Local Similarity 99.8%; Pred. No. 4.1e-187;
Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   7335 gttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaa 7394
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1938 gttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaa 1997

Qy   7395 caggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgct 7454
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1998 caggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgct 2057

Qy   7455 cctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacg 7514
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2058 cctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacg 2117

Qy   7515 agtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagg 7574
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2118 agtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagg 2177

Qy   7575 agaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtg 7634
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2178 agaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtg 2237

Qy   7635 gggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaag 7694
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2238 gggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaag 2297

Qy   7695 ctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtagaaggtcaca 7754
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   2298 ctctaaatgcccccgaattccagaccctgctggctgccagtaaataggtagaaggtcaca 2357

Qy   7755 cttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttga 7814
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2358 cttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttga 2417

Qy   7815 ctaatcgcttatctggattctgtaaaac 7842
          |||||||||||||||||||||||||||
Db   2418 ctaatcgcttatctggattctgtaaaac 2445
```

RESULT 5
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965

```
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1


  Query Match             4.6%;  Score 375;  DB 4;  Length 8060;
  Best Local Similarity   98.7%;  Pred. No. 6.7e-152;
  Matches 1395;  Conservative    0;  Mismatches   18;  Indels    1;  Gaps    1;

Qy   2973 gatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcgaatgg 3032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5495 GATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAACCACTTTTTGCCTTCGAATGG 5554

Qy   3033 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 3092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5555 AGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTC 5614

Qy   3093 aagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatc 3152
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   5615 AAGAACTCCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATC 5674

Qy   3153 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 3212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5675 CAACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACC 5734

Qy   3213 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 3272
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   5735 AAACAGGACTGCTTAGAAGGCACGAAGGCACTACTGCTGGAATTGTCTGACCTAGGCTAC 5794

Qy   3273 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 3332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5795 AGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGGTAACATACTTGGGGTACAGT 5854

Qy   3333 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 3392
          || |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5855 TTACGGGACGGGCAGCGATGGCTGACGGAGGCACGGAAGAAAACTGTAGTCCAGATACCG 5914

Qy   3393 gccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagactg 3452
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   5915 GCCCCAACCACAGCCAAACAAATGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTG 5974

Qy   3453 tggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 3512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5975 TGGATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGG 6034

Qy   3513 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 3572
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6035 GAATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATGCTATCAAAAAGGCCCTGCTG 6094

Qy   3573 agcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgag 3632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6095 AGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCTTTACCCTTTATGTGGATGAG 6154

Qy   3633 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 3692
          |||||||||||||||| |||||||||||||||||||||||||||||||| ||||||||
Db   6155 CGTAAGGGAGTAGCCCGGGGAGTTTTAACCCAAACCCTAGGACCATGGAGAAGACCTGTC 6214

Qy   3693 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 3752
          |||||||||||||||||||| |||||||||||||||||||||||| |||| |||||||||
Db   6215 GCCTACCTGTCAAAGAAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATGCCTGAAGGCT 6274

Qy   3753 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggac-agaatat 3811
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   6275 ATCGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAAGAATAT 6334

Qy   3812 aactgtaatagccccccatgcattggagaacatcgttcggcagcccccagaccgatggat 3871
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6335 AACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAGACCGATGGAT 6394

Qy   3872 gaccaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgc 3931
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   6395 GACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCACAGAGAGGGTCACGTTCGC 6454

Qy   3932 tccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgac 3991
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6455 TCCACCAACCGCTCTCAACCCTGCCACTCTTCTGCCTGAAGAGACTGATGAACCAGTGAC 6514

Qy   3992 tcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacat 4051
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6515 TCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACAT 6574

Qy   4052 accgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaagg 4111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6575 ACCGCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGG 6634

Qy   4112 taagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcct 4171
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   6635 TAAGAGGATGGCTGGGGCGGCGGTGGTGGACGGGACCCGCACGATCTGGGCCAGCAGCCT 6694

Qy   4172 gccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggct 4231
          ||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   6695 GCCGGGAGGAACTTCAGCACAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCT 6754

Qy   4232 ggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcaca 4291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6755 GGCCGAAGGGAAATCCATAAACATTTATACGGACAGCAGGTATGCCTTTGCGACTGCACA 6814

Qy   4292 cgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaa 4351
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6815 CGTACATGGGGCCATCTATAAACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAA 6874

Qy   4352 gaacaaagaggaaattctaagcctattagaagcc 4385
          |||||||||||||||||||||||||||||||||
Db   6875 GAACAAAGAGGAAATTCTAAGCCTATTAGAAGCC 6908


RESULT   6
US-09-376-781-10
; Sequence 10, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 278
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: sequence of part of PoEV env gene corresponding to
;   OTHER INFORMATION: residues 6057-6333 of WO 97/40167 for comparison.
US-09-376-781-10


  Query Match             3.4%;  Score 278;  DB 4;  Length 278;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-110;
```

```
  Matches  278;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   6057 gccgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagta 6116
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 gccgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagta 60

Qy   6117 caaaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaa 6176
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 caaaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaa 120

Qy   6177 gataagtttcactgaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagct 6236
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gataagtttcactgaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagct 180

Qy   6237 ggggaatagttttttattatatggcgggggagcagggtccactttaaccattcgccttag 6296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ggggaatagttttttattatatggcgggggagcagggtccactttaaccattcgccttag 240

Qy   6297 gatagagacggggacagaaccccctgtggcaatgggac 6334
          |||||||||||||||||||||||||||||||||||||
Db    241 gatagagacggggacagaaccccctgtggcaatgggac 278


RESULT   7
US-09-376-781-8
; Sequence 8, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 276
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: sequence of part of PERV-B env gene corresponding
;   OTHER INFORMATION: to residues 1350-1623 of GenBank Accession No. 6261806
;   OTHER INFORMATION: Y12239.
US-09-376-781-8


  Query Match             1.0%;  Score 86;  DB 4;  Length 276;
  Best Local Similarity   99.3%;  Pred. No. 1.4e-27;
  Matches  136;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   6057 gccgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagta 6116
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 gccgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagta 60

Qy   6117 caaaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaa 6176
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 caaagtgatgaaactatataaagataagagctgctccccatcagacttagattatctaaa 120

Qy   6177 gataagtttcactgaaa 6193
          |||||||||||||||||
Db    121 gataagtttcactgaaa 137


RESULT   8
US-09-075-272-1
; Sequence 1, Application US/09075272
; Patent No. 6136598
; GENERAL INFORMATION:
;   APPLICANT:  MILLER, A. DUSTY
;   APPLICANT:  WOLGAMOT, GREG
;   APPLICANT:  BONHAM, LYNN
;   TITLE OF INVENTION:  MUS DUNNI ENDOGENOUS RETROVIRAL
;   TITLE OF INVENTION:  PACKAGING CELL LINES
;   NUMBER OF SEQUENCES:  11
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;     STREET:  Two Embarcadero Center, 8th Floor
;     CITY:  San Francisco
;     STATE:  California
;     COUNTRY:  U.S.A.
;     ZIP:  94111
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/075,272
;     FILING DATE:  08-MAY-1998
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 60/046,140
;     FILING DATE:  09-MAY-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Poor, Brian W.
;     REGISTRATION NUMBER:  32,928
;     REFERENCE/DOCKET NUMBER:  14538A-003710
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (206) 467-9600
;     TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  8655 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
US-09-075-272-1


  Query Match             0.4%;  Score 35;  DB 3;  Length 8655;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-06;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5154 caaagctcaggacaggtagagaggatgaatagaac 5188
          |||||||||||||||||||||||||||||||||||
Db   5249 CAAAGCTCAGGACAGGTAGAGAGGATGAATAGAAC 5283


RESULT   9
US-08-798-000-3
; Sequence 3, Application US/08798000
; Patent No. 5766945
; GENERAL INFORMATION:
;   APPLICANT:  Miller, A. Dusty
;   TITLE OF INVENTION:  10A1 RETROVIRAL PACKAGING CELL LINES
;   NUMBER OF SEQUENCES:  3
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Townsend and Townsend and Crew LLP
;     STREET:  Two Embarcadero Center, 8th Floor
;     CITY:  San Francisco
;     STATE:  CA
;     COUNTRY:  USA
;     ZIP:  94111-3834
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/798,000
;      FILING DATE:  12-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/011,564
;      FILING DATE:  13-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  14538A-002010
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-467-9600
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  735 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-798-000-3


  Query Match             0.4%;  Score 29;  DB 1;  Length 735;
  Best Local Similarity   100.0%;  Pred. No. 0.0042;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5136 cattgtgcatacagaccccaaagctcagg 5164
          |||||||||||||||||||||||||||||
Db     96 CATTGTGCATACAGACCCCAAAGCTCAGG 124


RESULT  10
US-08-798-000-1
; Sequence 1, Application US/08798000
; Patent No. 5766945
;  GENERAL INFORMATION:
;    APPLICANT:  Miller, A. Dusty
;    TITLE OF INVENTION:  10A1 RETROVIRAL PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/798,000
;      FILING DATE:  12-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/011,564
;      FILING DATE:  13-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  14538A-002010
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-467-9600
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1234 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-798-000-1


  Query Match             0.4%;  Score 29;  DB 1;  Length 1234;
  Best Local Similarity   100.0%;  Pred. No. 0.0041;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5136 cattgtgcatacagaccccaaagctcagg 5164
          |||||||||||||||||||||||||||||
Db     96 CATTGTGCATACAGACCCCAAAGCTCAGG 124


RESULT  11
US-09-011-745-3
; Sequence 3, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 7308
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
US-09-011-745-3


  Query Match             0.4%;  Score 29;  DB 4;  Length 7308;
  Best Local Similarity   100.0%;  Pred. No. 0.0036;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   5136 cattgtgcatacagaccccaaagctcagg 5164
          |||||||||||||||||||||||||||||
Db   6178 cattgtgcatacagaccccaaagctcagg 6206


RESULT  12
US-09-011-745-4
; Sequence 4, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 7308
;   TYPE: DNA
```

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
US-09-011-745-4


  Query Match             0.4%;  Score 29;  DB 4;  Length 7308;
  Best Local Similarity   100.0%;  Pred. No. 0.0036;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5136 cattgtgcatacagaccccaaagctcagg 5164
           ||||||||||||||||||||||||||||||
Db    6178 cattgtgcatacagaccccaaagctcagg 6206


RESULT  13
US-09-011-745-2
; Sequence 2, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 7616
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
US-09-011-745-2


  Query Match             0.4%;  Score 29;  DB 4;  Length 7616;
  Best Local Similarity   100.0%;  Pred. No. 0.0036;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5136 cattgtgcatacagaccccaaagctcagg 5164
           ||||||||||||||||||||||||||||||
Db    5339 cattgtgcatacagaccccaaagctcagg 5367


RESULT  14
US-08-258-420-13/c
; Sequence 13, Application US/08258420
; Patent No. 5710037
;  GENERAL INFORMATION:
;    APPLICANT:  Nienhuis, Arthur W.
;    APPLICANT:  Vanin, Elio F.
;    TITLE OF INVENTION:  No. 5710037el Retroviral Envelope and LTR and Retroviral Vecto
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan,Cecchi, Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch diskette
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,420
;      FILING DATE:  10-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Olstein, Elliot M.
;      REGISTRATION NUMBER:  24,025
;      REFERENCE/DOCKET NUMBER:  271010-208
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8202 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  viral genome
US-08-258-420-13


  Query Match             0.4%;  Score 29;  DB 1;  Length 8202;
  Best Local Similarity   100.0%;  Pred. No. 0.0036;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    5136 cattgtgcatacagaccccaaagctcagg 5164
           ||||||||||||||||||||||||||||||
Db    2535 CATTGTGCATACAGACCCCAAAGCTCAGG 2507


RESULT  15
US-08-850-961-1
; Sequence 1, Application US/08850961
; Patent No. 6013517
;  GENERAL INFORMATION:
;    APPLICANT:  Respess, James G.
;    APPLICANT:  De Polo, Nicholas J.
;    APPLICANT:  Chada, Sunil
;    APPLICANT:  Sauter, Sybille
;    APPLICANT:  Bodner, Mordechai
;    APPLICANT:  Driver, David A.
;    TITLE OF INVENTION:  CROSSLESS RETROVIRAL VECTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation, Intellectual Property - R440
;      STREET:  P.O. Box 8097
;      CITY:  Emeryville
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/850,961
;      FILING DATE:  05-MAY-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kruse, No. 6013517man J.
;      REGISTRATION NUMBER:  35,235
;      REFERENCE/DOCKET NUMBER:  930049.424C4 / 1147.005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-3520
;      TELEFAX:  (510) 655-3542
```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-850-961-1

Query Match 0.4%; Score 29; DB 3; Length 8332;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    5136 cattgtgcatacagaccccaaagctcagg 5164
           ||||||||||||||||||||||||||||||
Db    5232 CATTGTGCATACAGACCCCAAAGCTCAGG 5260
```

Search completed: February 24, 2002, 07:30:07
Job time: 16416 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:30:07 ; Search time 197.25 Seconds
(without alignments)
3394.012 Million cell updates/sec

Title: US-09-171-553B-9
Perfect score: 2956
Sequence: 1 tgctttttagggttaggaac.....aaaaaaaaaaaaaaaaaaaa 2956

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1743	59.0	3482	3	US-09-111-085-3	Sequence 3, Appli
2	813	27.5	7333	4	US-08-766-528-2	Sequence 2, Appli
3	457	15.5	2462	3	US-09-111-085-1	Sequence 1, Appli
4	147	5.0	276	4	US-09-376-781-8	Sequence 8, Appli
5	137	4.6	278	4	US-09-376-781-10	Sequence 10, Appl
6	98	3.3	8132	4	US-08-766-528-3	Sequence 3, Appli
7	91	3.1	8060	4	US-08-766-528-1	Sequence 1, Appli
8	28	0.9	1493	4	US-09-376-781-24	Sequence 24, Appl
9	28	0.9	1736	4	US-09-360-197-13	Sequence 13, Appl
10	28	0.9	2000	4	US-09-376-781-25	Sequence 25, Appl
11	28	0.9	2000	4	US-09-376-781-30	Sequence 30, Appl
12	27	0.9	1587	3	US-09-108-020-11	Sequence 11, Appl
c 13	26	0.9	843	2	US-08-831-570-2	Sequence 2, Appli
14	26	0.9	1098	3	US-09-248-335-35	Sequence 35, Appl
c 15	26	0.9	1275	3	US-08-725-532A-2	Sequence 2, Appli
16	26	0.9	4032	1	US-08-126-587C-8	Sequence 8, Appli
17	26	0.9	80595	4	US-09-078-294-3	Sequence 3, Appli
c 18	25	0.8	25	4	US-09-376-781-23	Sequence 23, Appl
19	25	0.8	698	4	US-09-376-781-18	Sequence 18, Appl
20	25	0.8	945	4	US-09-268-364-3	Sequence 3, Appli
21	25	0.8	1000	4	US-09-357-251-19	Sequence 19, Appl
22	25	0.8	1001	1	US-08-728-259A-10	Sequence 10, Appl
23	25	0.8	1001	2	US-08-473-486-10	Sequence 10, Appl
24	25	0.8	1026	2	US-08-713-000-6	Sequence 6, Appli
25	25	0.8	1026	2	US-08-975-316-6	Sequence 6, Appli
26	25	0.8	1026	4	US-09-211-710-6	Sequence 6, Appli
27	25	0.8	1055	1	US-08-702-344-18	Sequence 18, Appl
28	25	0.8	1066	2	US-08-605-106-3	Sequence 3, Appli
29	25	0.8	1622	4	US-09-334-601-9	Sequence 9, Appli
30	25	0.8	1737	4	US-09-416-050A-7	Sequence 7, Appli
31	25	0.8	1737	4	US-09-664-800-7	Sequence 7, Appli
32	25	0.8	1737	4	US-09-665-309-7	Sequence 7, Appli
33	25	0.8	1737	4	US-09-661-569-7	Sequence 7, Appli
34	25	0.8	1826	2	US-08-808-931-15	Sequence 15, Appl
35	25	0.8	1826	3	US-08-808-323-15	Sequence 15, Appl
36	25	0.8	1826	3	US-09-050-603A-15	Sequence 15, Appl
37	25	0.8	1826	3	US-09-102-420B-15	Sequence 15, Appl
38	25	0.8	1889	3	US-08-861-747-3	Sequence 3, Appli
39	25	0.8	1957	4	US-09-352-990-11	Sequence 11, Appl
40	25	0.8	2158	1	US-07-602-608-1	Sequence 1, Appli
41	25	0.8	2158	1	US-08-261-578-1	Sequence 1, Appli
42	25	0.8	2351	1	US-08-168-091A-1	Sequence 1, Appli
43	25	0.8	3212	4	US-08-697-954-1	Sequence 1, Appli
44	25	0.8	4337	3	US-09-187-049-1	Sequence 1, Appli
c 45	24	0.8	40	4	US-09-306-290-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

Query Match 59.0%; Score 1743; DB 3; Length 3482;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2913; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

```
Qy       1 tgctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgctct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     548 tgctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgctct 607

Qy      61 acgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     608 acgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 667

Qy     121 cccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 180
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     668 cccagcctctgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 727

Qy     181 agctccgggaggcctactcaggagagagagacttgcaagttccacatcgcttccaagttg 240
           |||||||||||||||||||||||  ||||||||||||||||||||||||||||||| |
Db     728 agctccgggaggcctactcaggagaaggagacttgcaagttccacatcgcttccaagtgg 787

Qy     241 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggac 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     788 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggcc 847

Qy     301 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     848 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 907
```

Qy 361 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 420
Db 908 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 967
Qy 421 ctgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactccc 480
Db 968 ctgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactccc 1027
Qy 481 caggccagtagtaaacgccttatagacagctcgaacccccatagacctttatcccctacc 540
Db 1028 caggccagtagtaaacgccttatagacagctcgaacccccatagacctttatcccttacc 1087
Qy 541 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 600
Db 1088 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 1147
Qy 601 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 660
Db 1148 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 1207
Qy 661 agcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaa 720
Db 1208 agcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaa 1267
Qy 721 gagaaatactgtgggggttctgggggaatccttctgtaggagatggagctgcgtcacctcc 780
Db 1268 gagaaatactgtgggggttctgggggaatccttctgtaggagatggagctgcgtcacctcc 1327
Qy 781 aacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtc 840
Db 1328 aacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtc 1387
Qy 841 aattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctcccca 900
Db 1388 aattccggcccgggcaagtacaaagtgatgaaactatataaagataagagctgctcccca 1447
Qy 901 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 960
Db 1448 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 1507
Qy 961 aagtggataaatggtatgagctggggaatagtttttttaataatatggcgggggagcaggg 1020
Db 1508 aagtggataaatggtatgagctggggaatagtttttttataaatatggcgggggagcaggg 1567
Qy 1021 tccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatggga 1080
Db 1568 tccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcagtggga 1627
Qy 1081 cccgataaagtactggctgaacaggggccccccggccctggagccaccgcataacttgccg 1140
Db 1628 cccgataaagtactggctgaacaggggccccccggccctggagccaccgcataacttgccg 1687
Qy 1141 gtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccact 1200
Db 1688 gtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacggtaccact 1747
Qy 1201 ggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacag 1260
Db 1748 ggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacag 1807
Qy 1261 agactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgcc 1320
Db 1808 agactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgcc 1867
Qy 1321 acttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaa 1380
Db 1868 acttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaa 1927
Qy 1381 gaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaat 1440
Db 1928 gaaggaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaat 1987
Qy 1441 aagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcc 1500
Db 1988 aagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcc 2047
Qy 1501 caccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtattta 1560
Db 2048 caccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtattta 2107
Qy 1561 gtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacc 1620
Db 2108 gtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacc 2167
Qy 1621 tcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtcccccgagtgtac 1680
Db 2168 tcagtcttcaaccaatccaaagatttctgtgtcatggtccaaatcgtcccccgagtgtac 2227
Qy 1681 taccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaaga 1740
Db 2228 taccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaaga 2287
Qy 1741 gaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtagga 1800
Db 2288 gaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtagga 2347
Qy 1801 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 1860
Db 2348 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 2407
Qy 1861 catgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagag 1920
Db 2408 catgcggccatgacagaagatctccgagccttagaggagtctgttagcaacctagaagag 2467
Qy 1921 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 1980
Db 2468 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 2527
Qy 1981 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 2040
Db 2528 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 2587
Qy 2041 tcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagg 2100
Db 2588 tcaggagccatcagagactccatgagcaagcttagagaaaggttagagaggcgtcgaagg 2647
Qy 2101 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 2160
Db 2648 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 2707
Qy 2161 accctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttggg 2220
Db 2708 accctgctttctgctctgacgggacccctagtagtcctgctcctgttacttacagttggg 2767
Qy 2221 ccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatc 2280
Db 2768 ccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatc 2827
Qy 2281 atggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctagcct 2340
Db 2828 atggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctagcct 2887
Qy 2341 tcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaaaa 2400
Db 2888 tcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaaaa 2947
Qy 2401 tgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgcccccgaatt 2460
Db 2948 tgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgcccccgaatt 3007
Qy 2461 acagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggc 2520
Db 3008 ccagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggc 3067
Qy 2521 ctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggatt 2580

```
Db   3068 ctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggatt 3127

Qy   2581 ctgtaaaactgactggcaccatagaagaattgattacacattgacagccctagtgaccta 2640
Db   3128 ctgtaaaaccgactggcaccatagaagaattgattacacattgacagccctagtgaccta 3187

Qy   2641 tctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagcta 2700
Db   3188 tctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagcta 3247

Qy   2701 ttttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattggtccatg 2760
Db   3248 ttttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattggtccacg 3307

Qy   2761 gagcgcgggctc-tcgatattttaaaatgattggtttgtgacgcacaggctttgttgtga 2819
Db   3308 gagcgcgggctcttcgatattttaaaatgattggtttgtgacgcacaggctttgttgtga 3367

Qy   2820 accccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcgtggt 2879
Db   3368 accccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcgtggt 3427

Qy   2880 gtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 2934
Db   3428 gtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 3482
```

RESULT 2
US-08-766-528-2
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-2

Query Match 27.5%; Score 813; DB 4; Length 7333;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 863; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   2082 gttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaa 2141
Db   6470 GTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAA 6529

Qy   2142 caggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgct 2201
Db   6530 CAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTGCT 6589

Qy   2202 cctgttacttacagttgggccttgcttaattaataggtttgttgccttttgttagagaacg 2261
Db   6590 CCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACG 6649

Qy   2262 agtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagg 2321
Db   6650 AGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTGAGCCAAGG 6709

Qy   2322 agaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtg 2381
Db   6710 AGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACTATTAACAAGACAAGAAGTG 6769

Qy   2382 gggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaag 2441
Db   6770 GGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTAATAAAAAG 6829

Qy   2442 ctctaaatgccccccgaattacagaccctgctggctgccagtaaataggtagaaggtcaca 2501
Db   6830 CTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACA 6889

Qy   2502 cttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttga 2561
Db   6890 CTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGA 6949

Qy   2562 ctaatcgcttatctggattctgtaaaactgactggcaccatagaagaattgattacacat 2621
Db   6950 CTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACCATAGAAGAATTGATTACACAT 7009

Qy   2622 tgacagccctagtgacctatctcaactgcaatctgtcactctgcccaggagcccacgcag 2681
Db   7010 TGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAG 7069

Qy   2682 atgcggacctccggagctattttaaaatgattggtccacggagcgcgggctctcgatatt 2741
Db   7070 ATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATT 7129

Qy   2742 ttaaaatgattggtccatggagcgcgggctctcgatattttaaaatgattggtttgtgac 2801
Db   7130 TTAAAATGATTGGTCCATGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTTTGTGAC 7189

Qy   2802 gcacaggctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgcagt 2861
Db   7190 GCACAGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGT 7249

Qy   2862 cctctacccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctct 2921
Db   7250 CCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCT 7309

Qy   2922 tgctgtttgcatcaaaaaaaaaaa 2945
Db   7310 TGCTGTTTGCATCAAAAAAAAAAA 7333
```

RESULT 3
US-09-111-085-1
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences

```
;  FILE REFERENCE: 4238/75168
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2462
;   TYPE: DNA
;   ORGANISM: Porcine retrovirus
US-09-111-085-1


  Query Match             15.5%;  Score 457;  DB 3;  Length 2462;
  Best Local Similarity   99.8%;  Pred. No. 4.6e-195;
  Matches  507;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    2082 gttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaa 2141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1938 gttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaa 1997

Qy    2142 caggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgct 2201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1998 caggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgct 2057

Qy    2202 cctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacg 2261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2058 cctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacg 2117

Qy    2262 agtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagg 2321
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2118 agtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagg 2177

Qy    2322 agaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtg 2381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2178 agaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtg 2237

Qy    2382 gggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaag 2441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2238 gggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaag 2297

Qy    2442 ctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtagaaggtcaca 2501
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    2298 ctctaaatgcccccgaattccagaccctgctggctgccagtaaataggtagaaggtcaca 2357

Qy    2502 cttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttga 2561
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2358 cttcctattgttccagggcctgctatcctggcctaagtaagataacaggaaatgagttga 2417

Qy    2562 ctaatcgcttatctggattctgtaaaac 2589
           |||||||||||||||||||||||||||
Db    2418 ctaatcgcttatctggattctgtaaaac 2445


RESULT   4
US-09-376-781-8
; Sequence 8, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 276
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: sequence of part of PERV-B env gene corresponding
;   OTHER INFORMATION: to residues 1350-1623 of GenBank Accession No. 6261806
;   OTHER INFORMATION: Y12239.
US-09-376-781-8


  Query Match             5.0%;  Score 147;  DB 4;  Length 276;
  Best Local Similarity   99.5%;  Pred. No. 1.7e-56;
  Matches  197;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     801 gccgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagta 860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gccgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagta 60

Qy     861 caaaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaa 920
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 caaagtgatgaaactatataaagataagagctgctccccatcagacttagattatctaaa 120

Qy     921 gataagtttcactgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatgag 980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gataagtttcactgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatgag 180

Qy     981 ctggggaatagtttttta 998
           ||||||||||||||||||
Db     181 ctggggaatagtttttta 198


RESULT   5
US-09-376-781-10
; Sequence 10, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 278
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: sequence of part of PoEV env gene corresponding to
;   OTHER INFORMATION: residues 6057-6333 of WO 97/40167 for comparison.
US-09-376-781-10


  Query Match             4.6%;  Score 137;  DB 4;  Length 278;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-52;
  Matches  137;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     801 gccgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagta 860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gccgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaagta 60

Qy     861 caaaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaa 920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 caaaatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaa 120
```

```
Qy     921 gataagtttcactgaaa 937
           |||||||||||||||||
Db     121 gataagtttcactgaaa 137


RESULT   6
US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8132 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-3


  Query Match             3.3%;  Score 98;  DB 4;  Length 8132;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-34;
  Matches   98;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2846 cactcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgctt 2905
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8035 CACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCTT 8094

Qy    2906 ggaataaaaatcctcttgctgtttgcatcaaaaaaaaa 2943
           |||||||||||||||||||||||||||||||||||||
Db    8095 GGAATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAA 8132


RESULT   7
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1


  Query Match             3.1%;  Score 91;  DB 4;  Length 8060;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-31;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2846 cactcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgctt 2905
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2501 CACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCTT 2560

Qy    2906 ggaataaaaatcctcttgctgtttgcatcaa 2936
           |||||||||||||||||||||||||||||||
Db    2561 GGAATAAAAATCCTCTTGCTGTTTGCATCAA 2591


RESULT   8
US-09-376-781-24
; Sequence 24, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
;   LENGTH: 1493
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: sequence of the 3' end of the PERV-D env region.
US-09-376-781-24
```

Query Match 0.9%; Score 28; DB 4; Length 1493;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2047 gccatcagagactccatgaacaagctta 2074
||||||||||||||||||||||||||||
Db 1135 gccatcagagactccatgaacaagctta 1162

RESULT 9
US-09-360-197-13
; Sequence 13, Application US/09360197
; Patent No. 6287859
; GENERAL INFORMATION:
; APPLICANT: Bassilana, Frederic
; APPLICANT: Lazdunski, Michel
; APPLICANT: Waldmann, Rainer
; APPLICANT: Deweille, Jan R.
; TITLE OF INVENTION: Human and Rat Families of Neuronal Acid-Sensitive
; TITLE OF INVENTION: Cationic Channels, Their Cloning and Applications
; FILE REFERENCE: 989.6706P
; CURRENT APPLICATION NUMBER: US/09/360,197
; CURRENT FILING DATE: 1997-07-23
; PRIOR APPLICATION NUMBER: 09/129,758
; PRIOR FILING DATE: 1998-08-05
; PRIOR APPLICATION NUMBER: 60/095,408
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-360-197-13

Query Match 0.9%; Score 28; DB 4; Length 1736;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2929 tgcatcaaaaaaaaaaaaaaaaaaaaa 2956
||||||||||||||||||||||||||||
Db 1709 tgcatcaaaaaaaaaaaaaaaaaaaaa 1736

RESULT 10
US-09-376-781-25
; Sequence 25, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Contiguous
; OTHER INFORMATION: compilation of SEQ ID NOs: 18, 22, and 24.
US-09-376-781-25

Query Match 0.9%; Score 28; DB 4; Length 2000;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2047 gccatcagagactccatgaacaagctta 2074
||||||||||||||||||||||||||||
Db 1654 gccatcagagactccatgaacaagctta 1681

RESULT 11
US-09-376-781-30
; Sequence 30, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
; OTHER INFORMATION: consensus sequence of PERV-D.
US-09-376-781-30

Query Match 0.9%; Score 28; DB 4; Length 2000;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2047 gccatcagagactccatgaacaagctta 2074
||||||||||||||||||||||||||||
Db 1654 gccatcagagactccatgaacaagctta 1681

RESULT 12
US-09-108-020-11
; Sequence 11, Application US/09108020A
; Patent No. 6143561
; GENERAL INFORMATION:
; APPLICANT: Randall, Douglas D.
; APPLICANT: Johnston, Mark L.
; APPLICANT: Miernyk, Jan A.
; APPLICANT: Luethy, Michael H.
; APPLICANT: Mooney, Brian P.
; TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND
; TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO
; TITLE OF INVENTION: ENHANCE POLYHYDROXYALKANOATE BIOSYNTHESIS IN PLANTS
; FILE REFERENCE: UMO 1482
; CURRENT APPLICATION NUMBER: US/09/108,020A
; CURRENT FILING DATE: 1998-06-30
; EARLIER APPLICATION NUMBER: 60/051,291
; EARLIER FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 60/055,255
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/076,544
; EARLIER FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11

```
;   LENGTH: 1587
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-108-020-11


  Query Match             0.9%;  Score 27;  DB 3;  Length 1587;
  Best Local Similarity   100.0%;  Pred. No. 0.0065;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2930 gcatcaaaaaaaaaaaaaaaaaaaaaa 2956
          |||||||||||||||||||||||||||
Db   1553 gcatcaaaaaaaaaaaaaaaaaaaaaa 1579


RESULT  13
US-08-831-570-2/c
; Sequence 2, Application US/08831570
; Patent No. 5959175
;  GENERAL INFORMATION:
;    APPLICANT:  Thomas, Terry L.
;    APPLICANT:  Nunberg, Andrew N.
;    APPLICANT:  Beremand, Phillip D.
;    TITLE OF INVENTION:  A SUNFLOWER ALBUMIN 5' REGULATORY REGION
;    TITLE OF INVENTION:  FOR THE MODIFICATION OF PLANT SEED LIPID
;    TITLE OF INVENTION:  COMPOSITION
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Scully, Scott, Murphy & Presser
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/831,570
;      FILING DATE:  09-APR-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DiGiglio, Frank S.
;      REGISTRATION NUMBER:  31,346
;      REFERENCE/DOCKET NUMBER:  10545
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (516) 742-4343
;      TELEFAX:  (516) 743-4366
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  843 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-831-570-2


  Query Match             0.9%;  Score 26;  DB 2;  Length 843;
  Best Local Similarity   100.0%;  Pred. No. 0.019;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2931 catcaaaaaaaaaaaaaaaaaaaaaa 2956
          ||||||||||||||||||||||||||
Db    305 CATCAAAAAAAAAAAAAAAAAAAAAA 280


RESULT  14
US-09-248-335-35
; Sequence 35, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
;  APPLICANT: MCGONIGLE, BRIAN
;  APPLICANT: O'KEEF, DANIEL
;  TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE: CL-1128-A
;  CURRENT APPLICATION NUMBER: US/09/248,335
;  CURRENT FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: 08/924,759
;  EARLIER FILING DATE: 1997-September-05
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 35
;   LENGTH: 1098
;   TYPE: DNA
;   ORGANISM: maize
US-09-248-335-35


  Query Match             0.9%;  Score 26;  DB 3;  Length 1098;
  Best Local Similarity   100.0%;  Pred. No. 0.019;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2931 catcaaaaaaaaaaaaaaaaaaaaaa 2956
          ||||||||||||||||||||||||||
Db   1006 catcaaaaaaaaaaaaaaaaaaaaaa 1031


RESULT  15
US-08-725-532A-2/c
; Sequence 2, Application US/08725532A
; Patent No. 6020179
;  GENERAL INFORMATION:
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN TYROSINE PHOSPHATASES
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/725,532A
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0130 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1275 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:  Consensus
US-08-725-532A-2
```

Query Match 0.9%; Score 26; DB 3; Length 1275;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   2931 catcaaaaaaaaaaaaaaaaaaaaaa 2956
          ||||||||||||||||||||||||||
Db     40 CATCAAAAAAAAAAAAAAAAAAAAAA 15
```

Search completed: February 24, 2002, 07:31:17
Job time: 16486 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 21:04:09 ; Search time 212.72 Seconds
(without alignments)
3534.726 Million cell updates/sec

Title: US-09-171-553B-1
Perfect score: 3320
Sequence: 1 gaattcgcggccgcgtcgac.....aaaaaaaaaaaaaaaaaaaa 3320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2915.2	87.8	8132	4	US-08-766-528-3	Sequence 3, Appli
	2	2420.6	72.9	8060	4	US-08-766-528-1	Sequence 1, Appli
	3	1771.8	53.4	7333	4	US-08-766-528-2	Sequence 2, Appli
	4	1175.8	35.4	8655	3	US-09-075-272-1	Sequence 1, Appli
	5	1131	34.1	3482	3	US-09-111-085-3	Sequence 3, Appli
	6	1088	32.8	8535	3	US-08-716-351A-1	Sequence 1, Appli
	7	933.6	28.1	8323	1	US-08-110-300A-8	Sequence 8, Appli
	8	933.6	28.1	8323	2	US-08-886-642-8	Sequence 8, Appli
	9	933.6	28.1	8323	5	PCT-US93-08041-8	Sequence 8, Appli
c	10	932	28.1	10367	1	US-08-110-300A-9	Sequence 9, Appli
c	11	932	28.1	10367	2	US-08-886-642-9	Sequence 9, Appli
c	12	932	28.1	10367	5	PCT-US93-08041-9	Sequence 9, Appli
	13	916.2	27.6	7308	4	US-09-011-745-3	Sequence 3, Appli
	14	916.2	27.6	7308	4	US-09-011-745-4	Sequence 4, Appli
	15	916.2	27.6	7616	4	US-09-011-745-2	Sequence 2, Appli
	16	916.2	27.6	8332	3	US-08-850-961-1	Sequence 1, Appli
c	17	910.8	27.4	8202	1	US-08-258-420-13	Sequence 13, Appl
	18	847.8	25.5	2462	3	US-09-111-085-1	Sequence 1, Appli
	19	624.6	18.8	6363	2	US-08-929-967-6	Sequence 6, Appli
	20	423.8	12.8	445	4	US-09-376-781-7	Sequence 7, Appli
	21	372.4	11.2	698	4	US-09-376-781-18	Sequence 18, Appl
	22	371.8	11.2	2000	4	US-09-376-781-25	Sequence 25, Appl
	23	371.8	11.2	2000	4	US-09-376-781-30	Sequence 30, Appl
	24	307.2	9.3	1140	4	US-09-603-185-5	Sequence 5, Appli
	25	305.2	9.2	6312	4	US-09-011-745-7	Sequence 7, Appli
	26	300.4	9.0	949	5	PCT-US91-08254-1	Sequence 1, Appli
c	27	300.4	9.0	949	5	PCT-US91-08254-2	Sequence 2, Appli
	28	233	7.0	735	1	US-08-798-000-3	Sequence 3, Appli
	29	226.6	6.8	1234	1	US-08-798-000-1	Sequence 1, Appli
	30	221	6.7	1237	1	US-08-798-000-2	Sequence 2, Appli
	31	218	6.6	3674	1	US-08-105-483-324	Sequence 324, App
	32	218	6.6	3674	1	US-08-709-209-324	Sequence 324, App
	33	218	6.6	3674	1	US-08-458-101-324	Sequence 324, App
	34	216	6.5	9388	4	US-08-991-789A-141	Sequence 141, App
	35	204	6.1	3646	4	US-08-991-789A-227	Sequence 227, App
	36	183.4	5.5	2518	4	US-09-011-745-1	Sequence 1, Appli
	37	165.6	5.0	2448	3	US-08-691-563C-53	Sequence 53, Appl
	38	149.8	4.5	1167	3	US-08-691-563C-61	Sequence 61, Appl
	39	141.2	4.3	2389	3	US-08-691-563C-52	Sequence 52, Appl
	40	134.4	4.0	1101	4	US-08-991-789A-3	Sequence 3, Appli
	41	119.2	3.6	1158	1	US-08-471-724-1	Sequence 1, Appli
	42	119.2	3.6	1158	2	US-08-471-969-1	Sequence 1, Appli
	43	119.2	3.6	1158	2	US-08-384-137-1	Sequence 1, Appli
	44	119.2	3.6	1158	2	US-08-470-006A-1	Sequence 1, Appli
	45	119.2	3.6	1158	3	US-08-691-563C-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-3

Query Match 87.8%; Score 2915.2; DB 4; Length 8132;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 3051; Conservative 0; Mismatches 173; Indels 12; Gaps 2;

Qy 22 gatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaatgg 81
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db   2970 GATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAACCACTTTTTGCCTTCGAATGG 3029

Qy     82 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3030 AGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTC 3089

Qy    142 aagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatc 201
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   3090 AAGAACTCCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATC 3149

Qy    202 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3150 CAACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACC 3209

Qy    262 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 321
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3210 AAACAGGACTGCTTAGAAGGCACGAAGGCACTACTGCTGGAATTGTCTGACCTAGGCTAC 3269

Qy    322 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3270 AGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGGTAACATACTTGGGGTACAGT 3329

Qy    382 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 441
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3330 TTGCGGGACGGGCAGCGATGGCTGACGGAGGCACGGAAGAAAACTGTAGTCCAGATACCG 3389

Qy    442 gccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagactg 501
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   3390 GCCCCAACCACAGCCAAACAAATGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTG 3449

Qy    502 tggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3450 TGGATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGG 3509

Qy    562 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 621
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3510 GAATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATGCTATCAAAAAGGCCCTGCTG 3569

Qy    622 agcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgag 681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3570 AGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCTTTACCCTTTATGTGGATGAG 3629

Qy    682 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 741
          |||||||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db   3630 CGTAAGGGAGTAGCCCGGGGAGTTTTAACCCAAACCCTAGGACCATGGAGAAGACCTGTC 3689

Qy    742 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 801
          |||||||||||||||||||| |||||||||||||||||||||||||| |||| |||||||||
Db   3690 GCCTACCTGTCAAAGAAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATGCCTGAAGGCT 3749

Qy    802 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatata 861
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3750 ATCGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAGAATATA 3809

Qy    862 actgtaatagccccccatgcattggagaacatcgttcggcagcccccagaccgatggatg 921
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3810 ACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAGACCGATGGATG 3869

Qy    922 accaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgct 981
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   3870 ACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCACAGAGAGGGTCACGTTCGCT 3929

Qy    982 ccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgact 1041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3930 CCACCAGCCGCTCTCAACCCTGCCACTCTTCTGCCTGAAGAGACTGATGAACCAGTGACT 3989

Qy   1042 catgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacata 1101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3990 CATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATA 4049

Qy   1102 ccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggt 1161
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4050 CCGCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGGT 4109

Qy   1162 aagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctg 1221
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   4110 AAGAGGATGGCTGGGGCGGCGGTGGTGGACGGGACCCGCACGATCTGGGCCAGCAGCCTG 4169

Qy   1222 ccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctg 1281
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   4170 CCGGAAGGAACTTCAGCACAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCTG 4229

Qy   1282 gccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcacac 1341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4230 GCCGAAGGGAAATCCATAAACATTTATACGGACAGCAGGTATGCCTTTGCGACTGCACAC 4289

Qy   1342 gtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaag 1401
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4290 GTACATGGGGCCATCTATAAACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAAG 4349

Qy   1402 aacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctatt 1461
          |||||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db   4350 AACAAAGAGGAAATTCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCTATT 4409

Qy   1462 atacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggct 1521
          |||||||||||||||||||||||||||||||||| ||||||||||||| ||||| ||||||||||
Db   4410 ATACACTGTCCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCAGATGGCT 4469

Qy   1522 gaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgccc 1581
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||| ||||
Db   4470 GACCGGGTTGCCAAGCAGGCAGCCCAGGGTGTTAACCTTCTGCCTATAATAGAAATGCCC 4529

Qy   1582 aaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagata 1641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4530 AAAGCCCCAGAACCCAGACGACAGTACACCCTAGAAGACTGGCAAGAGATAAAAAAGATA 4589

Qy   1642 gaccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaatcctg 1701
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   4590 GACCAGTTCTCTGAGACTCCGGAAGGGACCTGCTATACCTCAGATGGGAAGGAAATCCTG 4649

Qy   1702 ccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaact 1761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4650 CCCCACAAAGAAGGGTTAGAATATGTCCAACAGATACATCGTCTAACCCACCTAGGAACT 4709

Qy   1762 aaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtg 1821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4710 AAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGGCTACCAGGAGTG 4769

Qy   1822 gctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccaga 1881
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4770 GCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGGTTAATGCTAATCCTTCCAGA 4829

Qy   1882 atacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttc 1941
          || |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4830 ATGCCTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTC 4889

Qy   1942 actgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacaccttt 2001
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4890 ACTGAGGTAAAGCCGGCTAAATACGGAAACAAATACCTATTGGTTTTTGTAGACACCTTT 4949

Qy   2002 tcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaa 2061
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   4950 TCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAAAAA 5009

Qy   2062 atactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggt 2121
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5010 ATACTGGAAGAAATTTTTCCAAGATTTGGAATACCTAAGGTAATAGGGTCAGACAATGGT 5069

Qy   2122 ccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaa 2181
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5070 CCAGCTTTTGTTGCCCAGGTAAGTCAGGGACTGGCCAAGATATTGGGGATTGATTGGAAA 5129

Qy   2182 ctgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccatt 2241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5130 CTGCATTGTGCATACAGACCCCAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATT 5189
```

```
Qy   2242 aaagagacccttaccaaattgaccacagagactggcattaatgattggatggctctcctg 2301
          |||||||||||||| |||||||| | |||||||| |||||||||||||| |||||||||
Db   5190 AAAGAGACCCTTACTAAATTGACCGCGGAGACTGGCGTTAATGATTGGATAGCTCTCCTG 5249

Qy   2302 ccctttgtgctttttagggtgaggaacacccctggacagtttgggctgacccccctatgaa 2361
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   5250 CCCTTTGTGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAA 5309

Qy   2362 ttgctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctgatgtg 2421
          || |||||||||||||||||| |||| ||||||||| | || |||||||||||| |||
Db   5310 TTACTCTACGGGGGACCCCCCCCCATTGGTAGAAATTGCTTCTGTACATAGTGCTGACGTG 5369

Qy   2422 ctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcgagcg 2481
          |||||||||||||||||||||||||||||||||| || |||||||||| || ||||||
Db   5370 CTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGCACTTGAGTGGGTGAGACAACGAGCG 5429

Qy   2482 tggaagcagctccgggaggcctactc---aggaggagacttgcaagttccacatcgcttc 2538
          |||| ||| |||||||||||||||   |||||||||||||| | |||||||| |||
Db   5430 TGGAGGCAACTCCGGGAGGCCTACTCAGGAGGAGGAGACTTGCAGATCCCACATCGTTTC 5489

Qy   2539 caagttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtgg 2598
          ||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   5490 CAAGTGGGAGATTCAGTCTACGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGG 5549

Qy   2599 aagggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatcccc 2658
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   5550 AAGGGCCCTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCC 5609

Qy   2659 acctggatccatgcatcccacgttaagccggcgccacctcccgattcggggtggaaagcc 2718
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   5610 ACCTGGATCCATGCATCCCACGTTAAACCGGCGCCACCTCCCGATTCGGGGTGGAAAGCC 5669

Qy   2719 gaaaagactgaaaatccccttaagcttcgcctccatcgcgtggttccttactctgtcaat 2778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5670 GAAAAGACTGAAAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTCAAT 5729

Qy   2779 aactcctcaagttaatggtaaacgccttgtggacagcccgaactcccataaaccccttatc 2838
          |||  ||||   |||||||  ||| | |  |||||| ||||||||||||||||||||||
Db   5730 AACCTCTCAGACTAATGGTATGCGCATAGGAGACAGCCTGAACTCCCATAAACCCTTATC 5789

Qy   2839 tctcacctggttacttactgactccggtacaggtattaatattaacagcactcaagggga 2898
          |||||||||||| |||||||||||| |||||||||||||| |||| ||||||||||||
Db   5790 TCTCACCTGGTTAATTACTGACTCCGGCACAGGTATTAATATCAACAACACTCAAGGGGA 5849

Qy   2899 ggctcccttggggacctggtggcctgaattatatgtctgccttcgatcagtaatccctgg 2958
          |||||| || || |||||||||||||  |||| || |||||  ||||||| || ||| |
Db   5850 GGCTCCTTTAGGAACCTGGTGGCCTGATCTATACGTTTGCCTCAGATCAGTTATTCCTAG 5909

Qy   2959 tctcaatgaccaggccacacccccgatgtactccgtgcttacgggttttacgtttgccc 3018
          |||         | || |||||||  ||| | |||| |||| |||| ||||| ||||||||
Db   5910 TCT---------GACCTCACCCCCAGATATCCTCCATGCTCACGGATTTTATGTTTGCCC 5960

Qy   3019 aggacccccaaataatgaagaatattgtggaaatcctcaggatttcctttgcaagcaatg 3078
          |||||| |||||||||| | || |||| |||||||    |||||| |||| || |||||
Db   5961 AGGACCACCAAATAATGGAAAACATTGCGGAAATCCCAGAGATTTCTTTTGTAAACAATG 6020

Qy   3079 gagctgcataacttctaatgatgggaattggaaatggccagtctctcagcaagacagagt 3138
          || |||  |||| ||||||||||  |||||||||||||  |||||||| || || ||
Db   6021 GAACTGTGTAACCTCTAATGATGGATATTGGAAATGGCCAACCTCTCAGCAGGATAGGGT 6080

Qy   3139 aagttactcttttgttaacaatcctaccagttataatcaatttaattatggccatgggag 3198
          |||||  |||| ||| ||||    |||||| | |   |||||||||||          |||
Db   6081 AAGTTTTTCTTATGTCAACACCTATACCAGCTCTGGACAATTTAATTACCTGACCTGGAT 6140

Qy   3199 atggaaagattggcaacagcgggtacaaaaagatgtacgaaataagcaaataagct 3254
            | |  |   | |   || |    |   |||  ||    |    ||||||| |
Db   6141 TAGAACTGGAAGCCCCAAGTGCTCTCCTTCAGACCTAGATTACCTAAAAATAAGTT 6196

RESULT   2
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1


  Query Match           72.9%;  Score 2420.6;  DB 4;  Length 8060;
  Best Local Similarity 97.5%;  Pred. No. 0;
  Matches 2502;  Conservative    0;  Mismatches   59;  Indels    6;  Gaps    4;

Qy     22 gatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaatgg 81
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5495 GATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAACCACTTTTTGCCTTCGAATGG 5554

Qy     82 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5555 AGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTC 5614

Qy    142 aagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatc 201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5615 AAGAACTCCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATC 5674

Qy    202 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5675 CAACACCCTCAGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACC 5734

Qy    262 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 321
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   5735 AAACAGGACTGCTTAGAAGGCACGAAGGCACTACTGCTGGAATTGTCTGACCTAGGCTAC 5794

Qy    322 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5795 AGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGGTAACATACTTGGGGTACAGT 5854

Qy    382 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 441
          || |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5855 TTACGGGACGGGCAGCGATGGCTGACGGAGGCACGGAAGAAAACTGTAGTCCAGATACCG 5914

Qy    442 gccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagactg 501
```

```
Db   5915 GCCCCAACCACAGCCAAACAAATGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTG 5974

Qy    502 tggatcccgggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 561

Db   5975 TGGATCCCGGGGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGG 6034

Qy    562 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 621

Db   6035 GAATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATGCTATCAAAAAGGCCCTGCTG 6094

Qy    622 agcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgag 681

Db   6095 AGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCTTTACCCTTTATGTGGATGAG 6154

Qy    682 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 741

Db   6155 CGTAAGGGAGTAGCCCGGGGAGTTTTAACCCAAACCCTAGGACCATGGAGAAGACCTGTC 6214

Qy    742 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 801

Db   6215 GCCTACCTGTCAAAGAAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATGCCTGAAGGCT 6274

Qy    802 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggac-agaatat 860

Db   6275 ATCGCAGCTGTGGCCATACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAAGAATAT 6334

Qy    861 aactgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatggat 920

Db   6335 AACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAGACCGATGGAT 6394

Qy    921 gaccaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgc 980

Db   6395 GACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCACAGAGAGGGTCACGTTCGC 6454

Qy    981 tccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgac 1040

Db   6455 TCCACCAACCGCTCTCAACCCTGCCACTCTTCTGCCTGAAGAGACTGATGAACCAGTGAC 6514

Qy   1041 tcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacat 1100

Db   6515 TCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACAT 6574

Qy   1101 accgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaagg 1160

Db   6575 ACCGCTGACTGGAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGG 6634

Qy   1161 taagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcct 1220

Db   6635 TAAGAGGATGGCTGGGGCGGCGGTGGTGGACGGGACCCGCACGATCTGGGCCAGCAGCCT 6694

Qy   1221 gccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggct 1280

Db   6695 GCCGGGAGGAACTTCAGCACAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCT 6754

Qy   1281 ggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcaca 1340

Db   6755 GGCCGAAGGGAAATCCATAAACATTTATACGGACAGCAGGTATGCCTTTGCGACTGCACA 6814

Qy   1341 cgtacacggggccatctataaacaaaggggggttgcttacctcagcagggagggaaataaa 1400

Db   6815 CGTACATGGGGCCATCTATAAACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAA 6874

Qy   1401 gaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctat 1460

Db   6875 GAACAAAGAGGAAATTCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCTAT 6934

Qy   1461 tatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggc 1520

Db   6935 TATACACTGTCCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCAGATGGC 6994

Qy   1521 tgaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcc 1580

Db   6995 TGACCGGGTTGCCAAGCAGGCAGCCCAGGGTGTTAACCTTCTGCCTATAATAGAAATGCC 7054

Qy   1581 caaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagat 1640

Db   7055 CAAAGCCCCAGAACCCAGACGACAGTACACCCTAGAAGACTGGCAAGAGATAAAAAAGAT 7114

Qy   1641 agaccagttctctgagactccggagggggacctgctatacctcatatgggaaggaaatcct 1700

Db   7115 AGACCA-TTCTCTGAGACTCCGGAAGGGACCTGCTATACCTCAGATGGGAAGGAAATCCT 7173

Qy   1701 gccccacaaagaagggttagaatatgtccaac-agatacatcgtctaacccacctaggaa 1759

Db   7174 GCCCCACAAAGAAGGGTTAGAATATGTCCAACAAGATACATCGTCTAACCCACCTAGGAA 7233

Qy   1760 ctaaacacctgcagcagttggtcagaacatcccccttatcatgttctgaggctaccaggag 1819

Db   7234 CTAAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGGCTACCAGGAG 7293

Qy   1820 tggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttcca 1879

Db   7294 TGGCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGGTTAATGCTAATCCTTCCA 7353

Qy   1880 gaatacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggact 1939

Db   7354 GAATGCCTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACT 7413

Qy   1940 tcactgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacct 1999

Db   7414 TCACTGAGGTAAAGCCGGCTAAATATGGAAACAAATACCTATTGGTTTTTGTAGACACCT 7473

Qy   2000 tttcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaaga 2059

Db   7474 TTTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTAGCTAAAA 7533

Qy   2060 aaatactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatg 2119

Db   7534 AAATACTGGAAGAAATTTTTCCAAGATTTGGAATACCTAAGGTAATAGGGTCAGACAATG 7593

Qy   2120 gtccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattgga 2179

Db   7594 GTCCAGCTTTTGTTGCCCAGGTAAGTCAGGGACTGGCCAAGATATTGGGGATTGATTGGA 7653

Qy   2180 aactgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaacca 2239

Db   7654 AACTGCATTGTGCATACAGACCCCAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCA 7713

Qy   2240 ttaaagagacccttaccaaattgaccacagagactggcattaatgattggatggctctcc 2299

Db   7714 TTAAAGAGACCCTTACTAAATTGACCGCGGAGACTGGCGTTAATGATTGGATAGCTCTCC 7773

Qy   2300 tgccctttgtgctttttagggtgaggaacaccccctggacagtttgggctgaccccctatg 2359

Db   7774 TGCCCTTTGTGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATG 7833

Qy   2360 aattgctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctgatg 2419

Db   7834 AATTACTCTACGGGGGACCCCCCCCATTGGTAGAAATTGCTTCTGTACATAGTGCTGATG 7893

Qy   2420 tgctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcgag 2479

Db   7894 TGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGCACTTGAGTGGGTGAGACAACGAG 7953

Qy   2480 cgtggaagcagctccgggaggcctactc---aggaggagacttgcaagttccacatcgct 2536

Db   7954 CGTGGAGGCAACTCCGGGAGGCCTACTCAGGAGGAGGAGACTTGCAGATCCCACATCGTT 8013

Qy   2537 tccaagttggagattcagtctatgttagacgccaccgtgcaggaaac 2583

Db   8014 TCCAAGTGGGAGATTCAGTCTACGTTAGACGCCACCGTGCAGGAAAC 8060

RESULT   3
US-08-766-528-2
```

```
; Sequence 2, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-2


  Query Match             53.4%;  Score 1771.8;  DB 4;  Length 7333;
  Best Local Similarity   95.1%;  Pred. No. 0;
  Matches 1871;  Conservative    0;  Mismatches   17;  Indels   79;  Gaps    1;

Qy   1351 gccatctataaacaaaggggttgcttacctcagcagggagggaaataaagaacaaagag 1410
          || | | | | | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3504 GCTAACCAAAGAAAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAG 3563

Qy   1411 gaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgt 1470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3564 GAAATTCTAAGCCTATTAGAAGCCTTACATTTGCCAAAAAGGCTAGCTATTATACACTGT 3623

Qy   1471 cctggacatcagaaagccaaagatctcatatctagagggaaccagatggctgaccgggtt 1530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3624 CCTGGACATCAGAAAGCCAAAGATCTCATATCTAGAGGGAACCAGATGGCTGACCGGGTT 3683

Qy   1531 gccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaaagccccca 1590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3684 GCCAAGCAGGCAGCCCAGGCTGTTAACCTTCTGCCTATAATAGAAACGCCCAAAGCCCCA 3743

Qy   1591 gaacccagacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttc 1650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3744 GAACCCAGACGACAGTACACCCTAGAAGACTGGCAAGAGATAAAAAAGATAGACCAGTTC 3803

Qy   1651 tctgagactccggaggggacctgctatacctcatatgggaaggaaatcctgccccacaaa 1710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3804 TCTGAGACTCCGGAGGGGACCTGCTATACCTCATATGGGAAGGAAATCCTGCCCCACAAA 3863

Qy   1711 gaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaaacacctg 1770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3864 GAAGGGTTAGAATATGTCCAACAGATACATCGTCTAACCCACCTAGGAACTAAACACCTG 3923

Qy   1771 cagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctgactcg 1830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3924 CAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGGCTACCAGGAGTGGCTGACTCG 3983

Qy   1831 gtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacctcca 1890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3984 GTGGTCAAACATTGTGTGCCCTGCCAGCTGGTTAATGCTAATCCTTCCAGAATACCTCCA 4043

Qy   1891 ggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactgaggta 1950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4044 GGAAAGAGACTAAGGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTA 4103

Qy   1951 aagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcaggatgg 2010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4104 AAGCCGGCTAAATACGGAAACAAATATCTATTGGTTTTTGTAGACACCTTTTCAGGATGG 4163

Qy   2011 gtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatactggag 2070
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4164 GTAGAGGCTTATCCTACTAAAAAAGAGACTTCAACCGTGGTGGCTAAGAAAATACTGGAG 4223

Qy   2071 gaaatttttccaagatttggaaatacctaaggtaatagggtcagacaatggtccagctttc 2130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4224 GAAATTTTTCCAAGATTTGGAATACCTAAGGTAATAGGGTCAGACAATGGTCCAGCTTTC 4283

Qy   2131 gttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgcattgt 2190
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4284 GTTGCCCAGGTAAGTCAGGGACTGGCCAAGATATTGGGGATTGATTGAAAACTGCATTGT 4343

Qy   2191 gcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacc 2250
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4344 GCATACAGACCCCAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACC 4403

Qy   2251 cttaccaaattgaccacagagactggcattaatgattggatggctctcctgccctttgtg 2310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4404 CTTACCAAATTGACCACAGAGACTGGCATTAATGATTGGATGGCTCTCCTGCCCTTTGTG 4463

Qy   2311 ctttttagggtgaggaacacccctggacagtttgggctgacccctatgaattgctctac 2370
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4464 CTTTTTAGGGTGAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATAAATTGCTCTAC 4523

Qy   2371 gggggaccccccccgttggcagaaattgcctttgcacatagtgctgatgtgctgctttcc 2430
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4524 GGGGGACCCCCCCCGTTGGCAGAAATTGCCTTTGCACATAGTGCTGATGTGCTGCTTTCC 4583

Qy   2431 cagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcgagcgtggaagcag 2490
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4584 CAGCCTTTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAGGCAGCGAGCGTGGAAGCAG 4643

Qy   2491 ctccgggaggcctactcaggaggagacttgcaagttccacatcgcttccaagttggagat 2550
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4644 CTCCGGGAGGCCTACTCAGGAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTTGGAGAT 4703

Qy   2551 tcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggaccttat 2610
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4704 TCAGTCTATGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTAGAAGGGACCTTAT 4763

Qy   2611 ctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatccccacctggatccat 2670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4764 CTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGG---------------------- 4804

Qy   2671 gcatcccacgttaagccggcgccacctcccgattcggggtggaaagccgaaaagactgaa 2730
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4805 ---------------------------------------------------------- 4804

Qy   2731 aatccccttaagcttcgcctccatcgcgtggttccttactctgtcaataactcctcaagt 2790
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4805 AATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTCAATAACTCCTCAAGT 4864

Qy   2791 taatggtaaacgccttgtggacagcccgaactcccataaaccctatctctcacctggtt 2850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4865 TAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTGGTT 4924
```

```
Qy   2851 acttactgactccggtacaggtattaatattaacagcactcaaggggaggctcccttggg 2910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4925 ACTTACTGACTCCGGTACAGGTATTAATATTAACAGCACTCAAGGGGAGGCTCCCTTGGG 4984

Qy   2911 gacctggtggcctgaattatatgtctgccttcgatcagtaatccctggtctcaatgacca 2970
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4985 GACCTGGTGGCCTGAATTATATGTCTGCCTTCGATCAGTAATCCCTGGTCTCAATGACCA 5044

Qy   2971 ggccacacccccgatgtactccgtgcttacgggttttacgtttgcccaggacccccaaa 3030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5045 GGCCACACCCCCCGATGTACTCCGTGCTTACGGGTTTTACGTTTGCCCAGGACCCCCAAA 5104

Qy   3031 taatgaagaatattgtggaaatcctcaggatttcctttgcaagcaatggagctgcataac 3090
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   5105 TAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAATGGAGCTGCATAAC 5164

Qy   3091 ttctaatgatgggaattggaaatggccagtctctcagcaagacagagtaagttactcttt 3150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5165 TTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTCTTT 5224

Qy   3151 tgttaacaatcctaccagttataatcaatttaattatggccatgggagatggaaagattg 3210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5225 TGTTAACAATCCTACCAGTTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGATTG 5284

Qy   3211 gcaacagcgggtacaaaaagatgtacgaaataagcaaataagctgtcattcgttagacct 3270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5285 GCAACAGCGGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTCATTCGTTAGACCT 5344

Qy   3271 agattacttaaaaataagtttcactaaaaaaaaaaaaaaaaaaaa 3317
          |||||||||||||||||||||||| |||||  |||| || |||| |
Db   5345 AGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATA 5391


RESULT   4
US-09-075-272-1
; Sequence 1, Application US/09075272
; Patent No. 6136598
;  GENERAL INFORMATION:
;    APPLICANT:  MILLER, A. DUSTY
;    APPLICANT:  WOLGAMOT, GREG
;    APPLICANT:  BONHAM, LYNN
;    TITLE OF INVENTION:  MUS DUNNI ENDOGENOUS RETROVIRAL
;    TITLE OF INVENTION:  PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,272
;      FILING DATE:  08-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/046,140
;      FILING DATE:  09-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poor, Brian W.
;      REGISTRATION NUMBER:  32,928
;      REFERENCE/DOCKET NUMBER:  14538A-003710
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-075-272-1


  Query Match             35.4%;  Score 1175.8;  DB 3;  Length 8655;
  Best Local Similarity   66.1%;  Pred. No. 0;
  Matches 1778;  Conservative    0;  Mismatches  877;  Indels   36;  Gaps    4;

Qy     19 acagatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcgaa 78
          | |||||||||||| ||||||  |  | |||||||  || ||||  || ||||||||| |||
Db   3095 AAAGATGCCTTCTTTTGCCTGCGTCTGCACCCCAAGAGTCAACTTCTGTTTGCCTTTGAA 3154

Qy     79 tggagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaaggg 138
          ||||||||  ||||     |||   || || ||  | ||||||||  |||| || || |||
Db   3155 TGGAGAGACCCAGAGGGCGGACAGACTGGTCAATTAACCTGGACTAGACTACCACAGGGG 3214

Qy    139 ttcaagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcagg 198
          |||||  |||||||| ||| | ||||||||||||||  ||  ||||  | ||  |||  |
Db   3215 TTCAAAAACTCCCCTACCCTGTTTGACGAAGCCCTCCATCGGGATTTAGCACCCTTTCGC 3274

Qy    199 atccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagcc 258
          |||  |||| ||| | ||||| || |||||||| |||||  ||  |  | || |  |||
Db   3275 GCTCAAAACCCCCAGCTTACCCTACTGCAGTACGTAGATGATCTTTTAATCGCAGCCGCC 3334

Qy    259 accaaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggc 318
          | ||| ||   ||   | | || ||  ||  ||| ||   ||| ||   |  | ||
Db   3335 TCAAAAGAGCTATGTCAACAGGGGACTGAGAGACTCCTCACAGAACTGGGGAATTTGGGG 3394

Qy    319 tacagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtac 378
          ||  |||  || |||||  |||||  || |||||     |  |||||||  ||||||||||
Db   3395 TATCGAGTTTCGGCTAAAAAGGCACAAATTTGTCAAACTGAGGTAATCTACTTGGGGTAT 3454

Qy    379 agtttgcggggcgggcagcgatggctgacggaggcacggaagaaaaactgtagtccagata 438
          | ||||||  ||  |  |||||||  || || ||  |||||  || |||||  |   |||
Db   3455 ACTTTGCGAGGAGGTAAAAGATGGCTCACTGAAGCCCGGAAAAAGACTGTCATGATGATT 3514

Qy    439 ccggccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcaga 498
          ||  |  |||||||||  |    ||  ||  |  ||||||  |||||||  |||||  |||
Db   3515 CCACCGCCAACCACACCACGGCAGGTACGTGAGTTTCTGGGGACTGCTGGCTTTTGTAGA 3574

Qy    499 ctgtggatcccgggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaa 558
          || |||||  || ||  |||||||||  |||||||  ||  ||  ||  |  ||  |  |||  |
Db   3575 CTCTGGATTCCAGGCTTTGCGACCCTAGCAGCACCCCTATATCCCTTGACTAGGGAAGGA 3634

Qy    559 gggggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctg 618
          ||     |||       || |||||  |  ||  |||||  |||||||||   |  | ||
Db   3635 ATTCCCTTTGAATGGAAAGAAGAACACCAAAGAGCTTTTGAGGCTATCAAATCGTCTCTA 3694

Qy    619 ctgagcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggat 678
          |||  || |||||  || ||  | || ||| | |||||  ||||   ||| ||||||||
Db   3695 ATGACTGCCCCTGCGCTAGCATTACCAGACTTGACTAAGTCCTTCGTCCTATATGTGGAC 3754

Qy    679 gagcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacct 738
          ||| |   |||  ||||| |||| ||  | || ||| | ||||||||  |||| ||||||
Db   3755 GAGAGAGCGGGCATAGCCAGAGGGGTGCTGACACAAGCACTAGGACCCTGGAAGAGACCG 3814

Qy    739 gttgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaag 798
          || ||||| |||||||| || || ||||| || || ||||| ||||||  |||||||||
Db   3815 GTAGCCTATCTGTCAAAAAAACTGGATCCCGTTGCTAGTGGATGGCCCACATGTCTGAAA 3874

Qy    799 gctatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaat 858
          ||||| ||||| || ||| | ||| |||| || ||||| ||| ||||  |||||||| |
Db   3875 GCTATTGCAGCAGTAGCCCTGCTGATCAAAGATGCTGATAAACTGACAATGGGACAGCAG 3934

Qy    859 ataactgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatgg 918
          | |||||  | ||||| |||||  ||||||  || || |||||||| || ||  ||||||
Db   3935 GTGACTGTCGTGGCCCCTCATGCCCTGGAGAGTATTGTGCGGCAGCCACCTGATAGATGG 3994

Qy    919 atgaccaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttc 978
          ||||| || ||||| |||||| |||||||| || ||||| || |  ||  |||| || ||
```

```
Db   3995 ATGACTAATGCCCGGATGACACACTATCAGAGTCTGCTGCTAAATGACCGGGTAACCTTT 4054

Qy    979 gctccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtg 1038

Db   4055 GCCCCCCCTGCCATTCTCAACCCAGCCACCCTCCTCCCTCTAACGAATGATTCCGTCCCA 4114

Qy   1039 actcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagac 1098

Db   4115 GTACATCGATGTGCAGACATCCTGGCTGAAGAAATTGGGACCAGAAAAGACCTGACTGAC 4174

Qy   1099 ataccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaa 1158

Db   4175 CAACCCTGGCCTGG---AGCTCCTAGCTGGTACACGGATGGCAGCAGTTTTCTGATTGAG 4231

Qy   1159 ggtaagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagc 1218

Db   4232 GGAAAGCGAAGGGCTGGAGCTGCGGTGGTGGATGGAAAAAAGGTAATTTGGGCAAGTGCC 4291

Qy   1219 ctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcgg 1278

Db   4292 TTGCCTGAAGGAACTTCGGCACAGAAGGCTGAACTCATAGCGCTTACACAGGCCCTCCGA 4351

Qy   1279 ctggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgca 1338

Db   4352 GAGGCCGAAGGTAAGATCATTAATATTTACACTGACAGCCGCTATGCTTTTGCTACTGCA 4411

Qy   1339 cacgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaata 1398

Db   4412 CATATCCATGGGGCCATCTACAGGCAGCGAGGGTTATTGACTTCAGCCGGTAAAGACATC 4471

Qy   1399 aagaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagct 1458

Db   4472 AAAAACAAAGAAGAAATTCTGGCCCTGTTAGAAGCCATACATGCGCCCAAGAAGGTAGCC 4531

Qy   1459 attatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatg 1518

Db   4532 ATCATCCACTGTCCCGGTCACCAGAAAGGAGAGGACTTAGTGGCCAAGGGCAACCGGATG 4591

Qy   1519 gctgaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacg 1578

Db   4592 GCAGACTCAGTAGCAAAACAGGTTGCTCAGGGGGCCATGATCTTAACTGAAAAAGGAAAT 4651

Qy   1579 cccaaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaag 1638

Db   4652 CC----GTCCAAAAGCCCTGA----------------------GGATGAAAACTATGAT 4684

Qy   1639 atagaccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaatc 1698

Db   4685 ATAAAAGAACTATTTTGGACTAGTGATCCCCTCCCATACTTTTTCGAAGGAAAAATAGAC 4744

Qy   1699 ctgccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctagga 1758

Db   4745 TTGACACCCGAGGAAGGAATAAAATTTGTGAAAGGACTACACCAGTTTACCCACCTGGGA 4804

Qy   1759 actaaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccagga 1818

Db   4805 GTTGAAAAAATGATGAGACTGATTAAAAAGTCCCGGTATCAAGTCCCTAACTTGAAGTCA 4864

Qy   1819 gtggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttcc 1878

Db   4865 GTGGCTCAAAAGATTATAAACTCCTGCAAAGCGTGTGCATTCACTAATGCAACTAAAACC 4924

Qy   1879 agaatacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggac 1938

Db   4925 TACAAAGAACCTGGAAAGAGACAACGGGGAGACCGTCCTGGAGTGTATTGGGAGGTGGAT 4984

Qy   1939 ttcactgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacc 1998

Db   4985 TTTACTGAAGTTAAACCCGGAATGTATGGTAATAAGTATCTGTTAGTATTTGTAGACACC 5044

Qy   1999 ttttcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaag 2058

Db   5045 TTTTCAGGATGGGTTGAAGCGTTCCCCACTAAAACTGAGACTGCCCAGATTGTGGCCAAG 5104
```

```
Qy   2059 aaaatactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaat 2118

Db   5105 AAGATTTTTGAAGAAATCCTGCCAAGATATGGAGTACCTAAGGTAATCGGGTCCGACAAT 5164

Qy   2119 ggtccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattgg 2178

Db   5165 GGACCAGCCTTTGTTGCCCAGGTAAGTCAGGGATTGGCCACTCAGTTGGGCATTGATTGG 5224

Qy   2179 aaactgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaacc 2238

Db   5225 AAATTACATTGTGCTTACCGCCCTCAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACA 5284

Qy   2239 attaaagagacccttaccaaattgaccacagagactggcattaatgattggatggctctc 2298

Db   5285 TTAAAGGAGACCTTGACTAAATTGGCCATGGAGACCGGCGGGAAAGACTGGGTGGCTCTC 5344

Qy   2299 ctgccctttgtgcttttttagggtgaggaacacccctggacagtttgggctgacccccctat 2358

Db   5345 CTCCCCCTTGCGCTCTTCCGAGCCCGGAACACCCCCGGACGTTTTGGGCTCACTCCTTTT 5404

Qy   2359 gaattgctctacgggggaccccccccgttggcagaaattg------cctttgcacatagt 2412

Db   5405 GAAGTTCTGTATGGGGGACCTCCCCCTTTAATAAAAGATGGTGGAACATTGGTTCCCGAT 5464

Qy   2413 gctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgagg 2472

Db   5465 TCAGGCTCTGTCTTACCCTCCTCTTTGCTTATTCATTTAAAGGCCCTGAAAGTAATTAGG 5524

Qy   2473 cagcgagcgtggaagcagctccgggaggcctactcaggaggagacttgcaagttccacat 2532

Db   5525 ACCCAAATTTGGGACCAGCTAAAGACGGCCTACACCCCAGGGACCACCGCAGTACCCCAC 5584

Qy   2533 cgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagact 2592

Db   5585 GAGTTCCAGGTCGGGGACCAAGTCTTGGTCAGACGACATCGAACCGGTAGCCTTGAACCA 5644

Qy   2593 cggtggaagggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaagga 2652

Db   5645 CGGTGGAAGGGACCCTATTTAGTGTTACTAACAACTCCTACGGCAGTGAAAGTTGACGGG 5704

Qy   2653 atccccacctggatccatgcatcccacgttaagccggcgccacctcccgat 2703

Db   5705 ATTGCCTCCTGGATCCACGCTTCACACGTCAAGAGGGCCCCCAGTCAAGAT 5755
```

RESULT 5
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

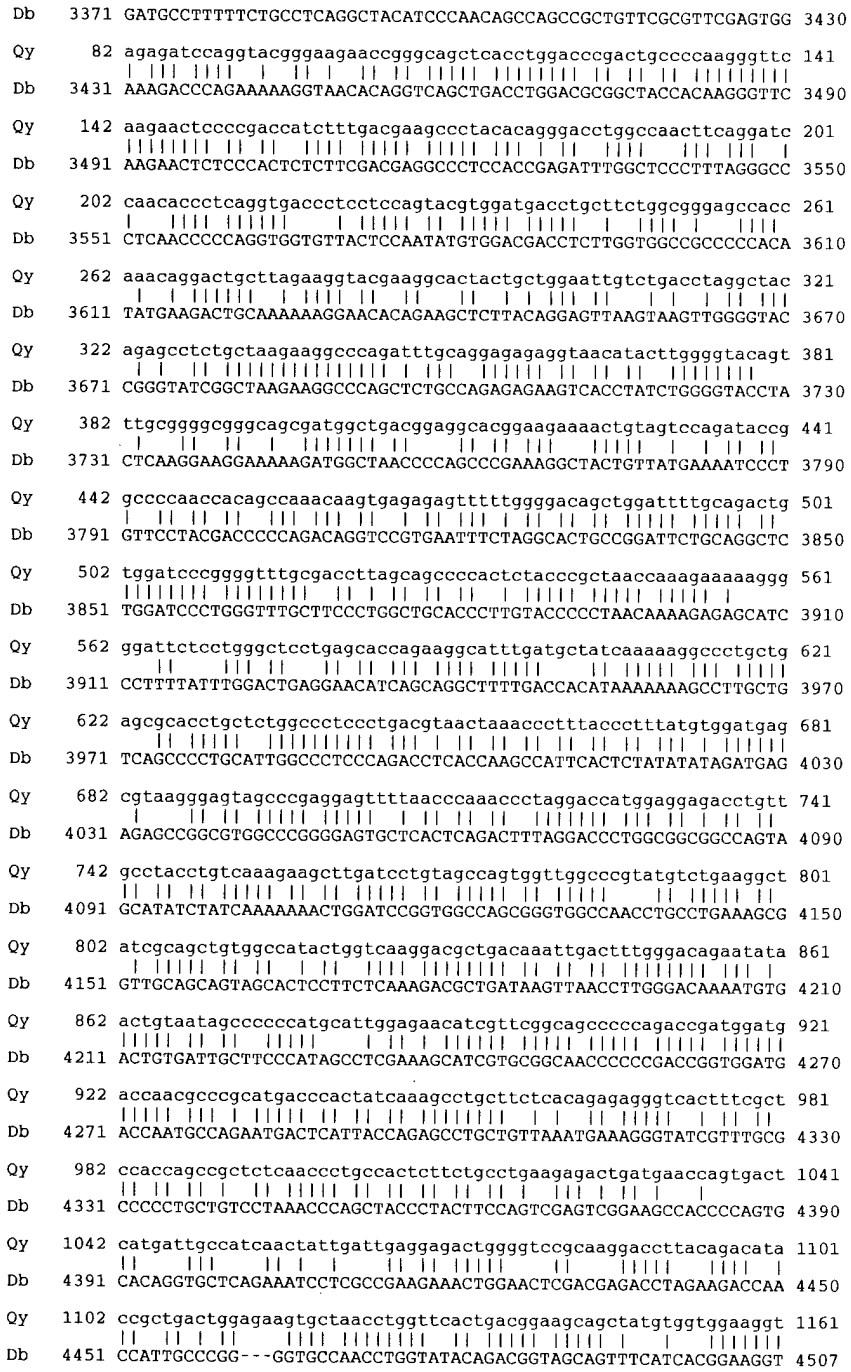
Query Match 34.1%; Score 1131; DB 3; Length 3482;
Best Local Similarity 89.8%; Pred. No. 0;
Matches 1251; Conservative 0; Mismatches 130; Indels 12; Gaps 3;

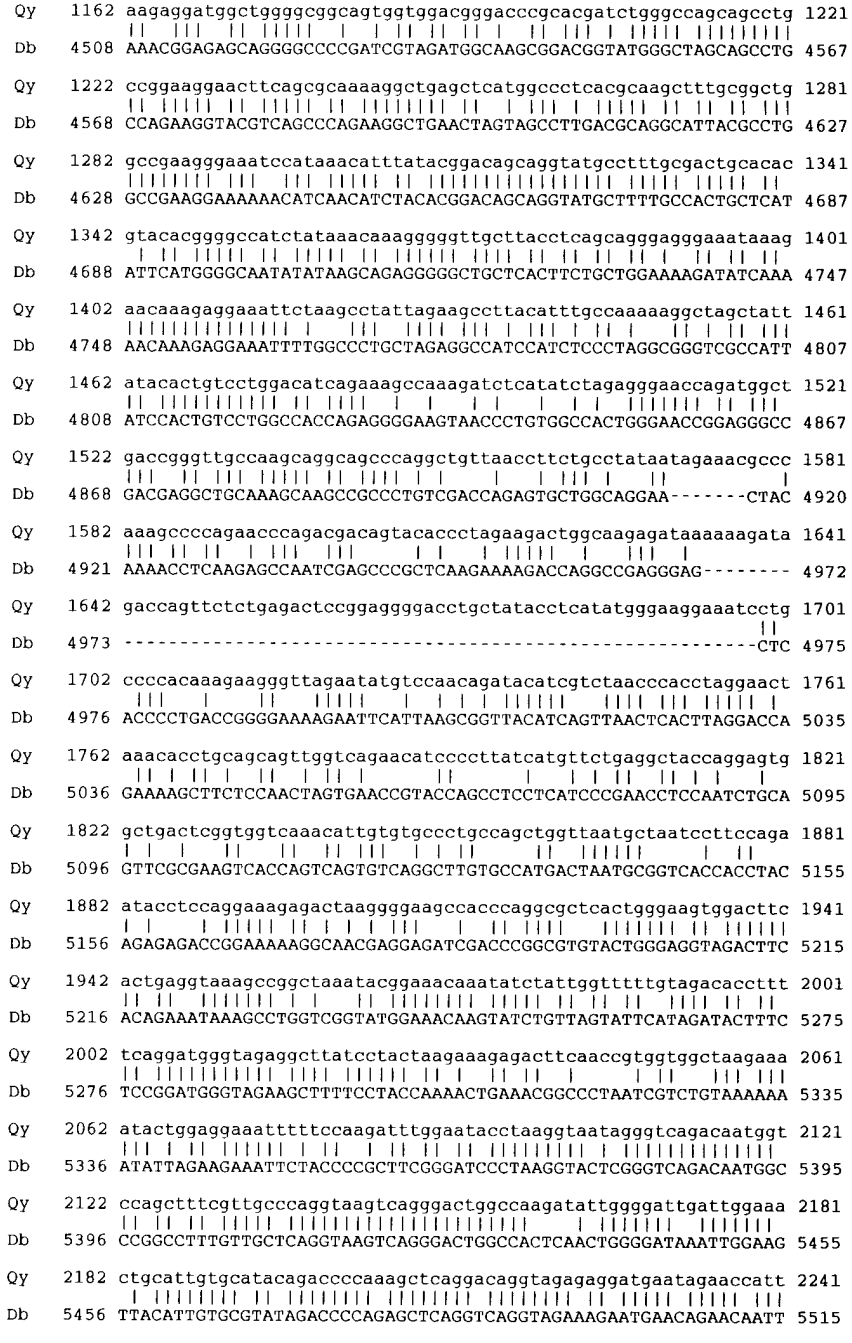
```
Qy   1767 cctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctga 1826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 cctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctga 65

Qy   1827 ctcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacc 1886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 ctcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacc 125

Qy   1887 tccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactga 1946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    126 tccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactga 185

Qy   1947 ggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcagg 2006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 ggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcagg 245

Qy   2007 atgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatact 2066
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 atgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaaaaaaatact 305

Qy   2067 ggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagc 2126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    306 ggaggaaatttttccgagatttggaatacctaaggtaatcgggtcagacaatggtccagc 365

Qy   2127 tttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgca 2186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 ttttgttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgca 425

Qy   2187 ttgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaaga 2246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 ttgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaaga 485

Qy   2247 gacccttaccaaattgaccacagagactggcattaatgattggatggctctcctgccctt 2306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    486 gacccttaccaaattgaccacagagactggcattaatgattggatagctctcctgccctt 545

Qy   2307 tgtgctttttagggtgaggaacacccctggacagtttgggctgacccccctatgaattgct 2366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    546 tgtgctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgct 605

Qy   2367 ctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctgatgtgctgct 2426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    606 ctacgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgct 665

Qy   2427 ttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcgagcgtggaa 2486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    666 ttcccagcctctgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaa 725

Qy   2487 gcagctccgggaggcctactcagg---aggagacttgcaagttccacatcgcttccaagt 2543
          |||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db    726 gcagctccgggaggcctactcaggagaaggagacttgcaagttccacatcgcttccaagt 785

Qy   2544 tggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaaggg 2603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    786 gggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaaggg 845

Qy   2604 accttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatccccacctg 2663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    846 cccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctg 905

Qy   2664 gatccatgcatcccacgttaagccggcgccacctcccgattcggggtggaaagccgaaaa 2723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    906 gatccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaa 965

Qy   2724 gactgaaaatccccttaagcttcgcctccatcgcgtggttccttactctgtcaataactc 2783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    966 gactgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactc 1025

Qy   2784 ctcaagttaatggtaaacgccttgtggacagcccgaactcccataaacccttatctctca 2843
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1026 cccaggccagtagtaaacgccttatagacagctcgaaccccatagacctttatccctta 1085

Qy   2844 cctggttacttactgactccggtacaggtattaatattaacagcactcaaggggaggctc 2903
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1086 cctggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctc 1145

Qy   2904 ccttggggacctggtggcctgaattatatgtctgccttcgatcagtaatccctggtctca 2963
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1146 ctagaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgtta 1205

Qy   2964 atgaccaggccacacccccgatgtactccgtgcttacgggttttacgtttgcccaggac 3023
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1206 ------aaagcacacctcccaacctagtccgtagttatgggttctattgctgcccagg-- 1257

Qy   3024 ccccaaataatgaagaatattgtggaaatcctcaggatttccttgcaagcaatggagct 3083
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1258 -cacagagaaagagaaatactgtgggggttctggggaatccttctgtaggagatggagct 1316

Qy   3084 gcataacttctaatgatgggaattggaaatggccagtctctcagcaagacagagtaagtt 3143
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1317 gcgtcacctccaacgatggagactggaaatggccgatctctctccaggaccgggtaaaat 1376

Qy   3144 actcttttgttaa 3156
          ||||||||||||
Db   1377 tctcctttgtcaa 1389
```

RESULT 6
US-08-716-351A-1
; Sequence 1, Application US/08716351A
; Patent No. 6033905
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Gibbon Ape Leukemia Virus-Based
; TITLE OF INVENTION: Retroviral Vectors
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,351A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/03784
; FILING DATE: 06-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian,, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 15280-128-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..8535
; OTHER INFORMATION: /standard_name= "GaLV SEATO Genome"
US-08-716-351A-1

Query Match 32.8%; Score 1088; DB 3; Length 8535;
Best Local Similarity 63.8%; Pred. No. 0;
Matches 1761; Conservative 0; Mismatches 920; Indels 81; Gaps 4;

```
Qy     22 gatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaatgg 81
          ||||||||| |||||||| ||  |||| ||||  ||||| || || || || ||||| |||
```

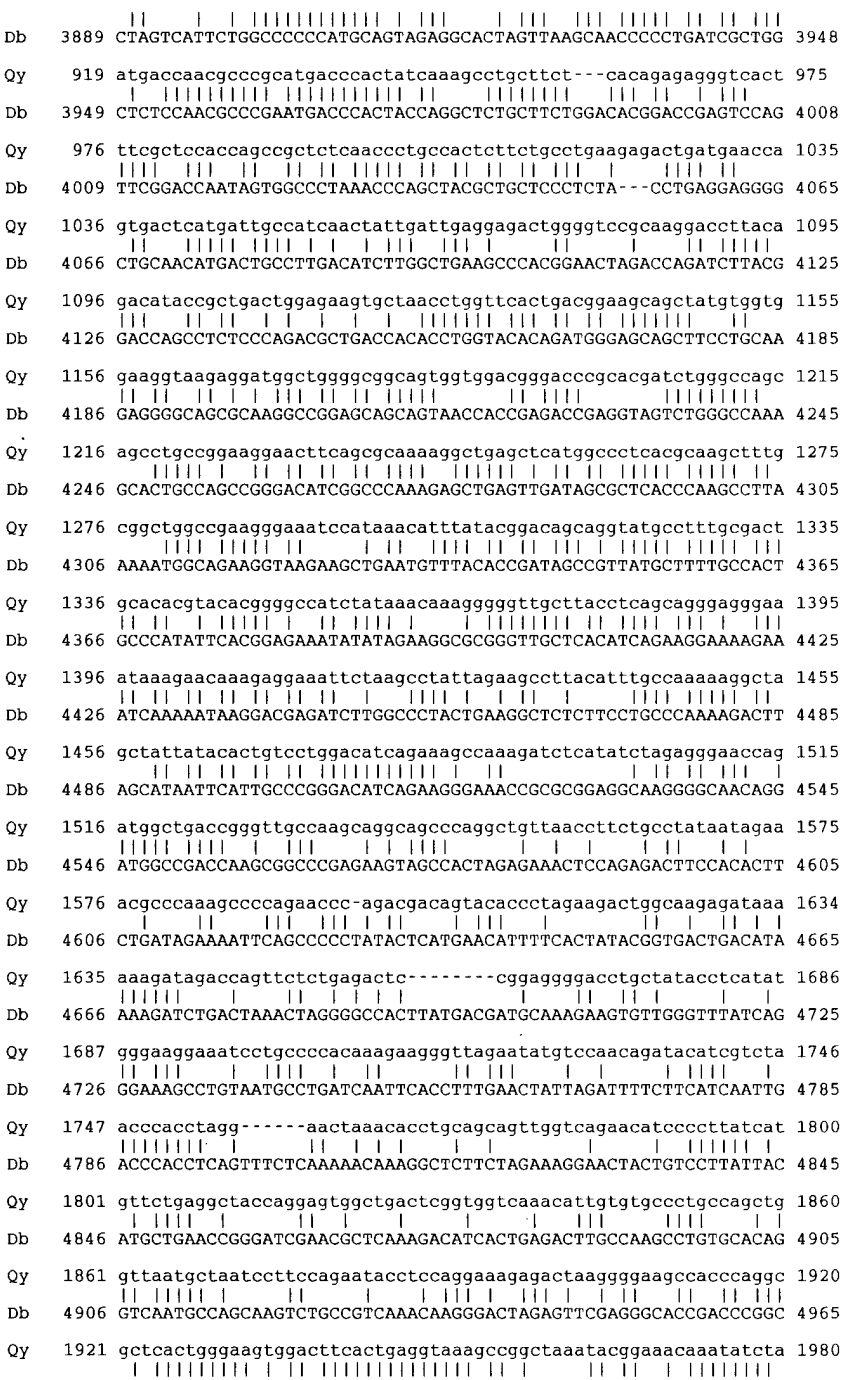

Db 3371 GATGCCTTTTTCTGCCTCAGGCTACATCCCAACAGCCAGCCGCTGTTCGCGTTCGAGTGG 3430
Qy 82 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 141
Db 3431 AAAGACCCAGAAAAAGGTAACACAGGTCAGCTGACCTGGACGCGGCTACCACAAGGGTTC 3490
Qy 142 aagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatc 201
Db 3491 AAGAACTCTCCCACTCTCTTCGACGAGGCCCTCCACCGAGATTTGGCTCCCTTTAGGGCC 3550
Qy 202 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 261
Db 3551 CTCAACCCCCAGGTGGTGTTACTCCAATATGTGGACGACCTCTTGGTGGCCGCCCCCACA 3610
Qy 262 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 321
Db 3611 TATGAAGACTGCAAAAAAGGAACACAGAAGCTCTTACAGGAGTTAAGTAAGTTGGGGTAC 3670
Qy 322 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 381
Db 3671 CGGGTATCGGCTAAGAAGGCCCAGCTCTGCCAGAGAGAAGTCACCTATCTGGGGTACCTA 3730
Qy 382 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 441
Db 3731 CTCAAGGAAGGAAAAAGATGGCTAACCCCAGCCCGAAAGGCTACTGTTATGAAAATCCCT 3790
Qy 442 gccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagactg 501
Db 3791 GTTCCTACGACCCCCAGACAGGTCCGTGAATTTCTAGGCACTGCCGGATTCTGCAGGCTC 3850
Qy 502 tggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 561
Db 3851 TGGATCCCTGGGTTTGCTTCCCTGGCTGCACCCTTGTACCCCCTAACAAAAGAGAGCATC 3910
Qy 562 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 621
Db 3911 CCTTTTATTTGGACTGAGGAACATCAGCAGGCTTTTGACCACATAAAAAAAGCCTTGCTG 3970
Qy 622 agcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgag 681
Db 3971 TCAGCCCCTGCATTGGCCCTCCCAGACCTCACCAAGCCATTCACTCTATATATAGATGAG 4030
Qy 682 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 741
Db 4031 AGAGCCGGCGTGGCCCGGGGAGTGCTCACTCAGACTTTAGGACCCTGGCGGCGGCCAGTA 4090
Qy 742 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 801
Db 4091 GCATATCTATCAAAAAAACTGGATCCGGTGGCCAGCGGGTGGCCAACCTGCCTGAAAGCG 4150
Qy 802 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatata 861
Db 4151 GTTGCAGCAGTAGCACTCCTTCTCAAAGACGCTGATAAGTTAACCTTGGGACAAAATGTG 4210
Qy 862 actgtaatagccccccatgcattggagaacatcgttcggcagcccccagaccgatggatg 921
Db 4211 ACTGTGATTGCTTCCCATAGCCTCGAAAGCATCGTGCGGCAACCCCCCGACCGGTGGATG 4270
Qy 922 accaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgct 981
Db 4271 ACCAATGCCAGAATGACTCATTACCAGAGCCTGCTGTTAAATGAAAGGGTATCGTTTGCG 4330
Qy 982 ccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgact 1041
Db 4331 CCCCCTGCTGTCCTAAACCCAGCTACCCTACTTCCAGTCGAGTCGGAAGCCACCCCAGTG 4390
Qy 1042 catgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagacata 1101
Db 4391 CACAGGTGCTCAGAAATCCTCGCCGAAGAAACTGGAACTCGACGAGACCTAGAAGACCAA 4450
Qy 1102 ccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggt 1161
Db 4451 CCATTGCCCGG---GGTGCCAACCTGGTATACAGACGGTAGCAGTTTCATCACGGAAGGT 4507


Qy 1162 aagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctg 1221
Db 4508 AAACGGAGAGCAGGGGCCCCGATCGTAGATGGCAAGCGGACGGTATGGGCTAGCAGCCTG 4567
Qy 1222 ccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctg 1281
Db 4568 CCAGAAGGTACGTCAGCCCAGAAGGCTGAACTAGTAGCCTTGACGCAGGCATTACGCCTG 4627
Qy 1282 gccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcacac 1341
Db 4628 GCCGAAGGAAAAAACATCAACATCTACACGGACAGCAGGTATGCTTTTGCCACTGCTCAT 4687
Qy 1342 gtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaataaag 1401
Db 4688 ATTCATGGGGCAATATATAAGCAGAGGGGGCTGCTCACTTCTGCTGGAAAAGATATCAAA 4747
Qy 1402 aacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagctatt 1461
Db 4748 AACAAAGAGGAAATTTTGGCCCTGCTAGAGGCCATCCATCTCCCTAGGCGGGTCGCCATT 4807
Qy 1462 atacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggct 1521
Db 4808 ATCCACTGTCCTGGCCACCAGAGGGGAAGTAACCCTGTGGCCACTGGGAACCGGAGGGCC 4867
Qy 1522 gaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaaacgccc 1581
Db 4868 GACGAGGCTGCAAAGCAAGCCGCCCTGTCGACCAGAGTGCTGGCAGGAA-------CTAC 4920
Qy 1582 aaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaagata 1641
Db 4921 AAAACCTCAAGAGCCAATCGAGCCCGCTCAAGAAAAGACCAGGCCGAGGGAG-------- 4972
Qy 1642 gaccagttctctgagactccggaggggacctgctatacctcatatgggaaggaaatcctg 1701
Db 4973 ----------CTC 4975
Qy 1702 ccccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaact 1761
Db 4976 ACCCCTGACCGGGGAAAAGAATTCATTAAGCGGTTACATCAGTTAACTCACTTAGGACCA 5035
Qy 1762 aaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtg 1821
Db 5036 GAAAAGCTTCTCCAACTAGTGAACCGTACCAGCCTCCTCATCCCGAACCTCCAATCTGCA 5095
Qy 1822 gctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccaga 1881
Db 5096 GTTCGCGAAGTCACCAGTCAGTGTCAGGCTTGTGCCATGACTAATGCGGTCACCACCTAC 5155
Qy 1882 atacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttc 1941
Db 5156 AGAGAGACCGGAAAAAGGCAACGAGGAGATCGACCCGGCGTGTACTGGGAGGTAGACTTC 5215
Qy 1942 actgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacaccttt 2001
Db 5216 ACAGAAATAAAGCCTGGTCGGTATGGAAACAAGTATCTGTTAGTATTCATAGATACTTTC 5275
Qy 2002 tcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaa 2061
Db 5276 TCCGGATGGGTAGAAGCTTTTCCTACCAAAACTGAAACGGCCCTAATCGTCTGTAAAAAA 5335
Qy 2062 atactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggt 2121
Db 5336 ATATTAGAAGAAATTCTACCCCGCTTCGGGATCCCTAAGGTACTCGGGTCAGACAATGGC 5395
Qy 2122 ccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaa 2181
Db 5396 CCGGCCTTTGTTGCTCAGGTAAGTCAGGGACTGGCCACTCAACTGGGGATAAATTGGAAG 5455
Qy 2182 ctgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccatt 2241
Db 5456 TTACATTGTGCGTATAGACCCCAGAGCTCAGGTCAGGTAGAAAGAATGAACAGAACAATT 5515

```
Qy   2242 aaagagacccttaccaaattgaccacagagactggcattaatgattggatggctctcctg 2301
          |||||||||| | ||||||||  ||  ||||||  ||    || || ||| || | ||||||
Db   5516 AAAGAGACCTTGACCAAATTAGCCTTAGAGACCGGTGGAAAAGACTGGGTGACCCTCCTT 5575

Qy   2302 ccctttgtgctttttagggtgaggaacacccctggacagtttgggctgacccccctatgaa 2361
          ||||| | |||  ||||||  ||||| ||||||||  | ||||||  | || || ||||||
Db   5576 CCCTTAGCGCTGCTTAGGGCCAGGAATACCCCTGGCCGGTTTGGTTTAACTCCTTATGAA 5635

Qy   2362 ttgctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctgatgtg 2421
           | ||||| || ||||| |||||  |    ||   ||       |    || | ||||
Db   5636 ATTCTCTATGGAGGACCACCCCCCATACTTGAGTCTGGAGAAACTTTGGGTCCCGATGAT 5695

Qy   2422 ctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcgagcg 2481
               ||  ||      || ||  ||     | |||||  | ||    || |||    | |
Db   5696 AGATTTCTCCCTGTCTTATTTACTCACTTAAAGGCTTTAGAAATTGTAAGGACCCAAATC 5755

Qy   2482 tggaagcagctccgggaggcctactcaggaggagacttgcaagttccacatcgcttccaa 2541
          ||| | ||| ||   ||||  ||       ||   | |   | | || || |  |||||
Db   5756 TGGGACCAGATCAAAGAGGTGTATAAGCCTGGTACCGTAACAATCCCTCACCCGTTCCAG 5815

Qy   2542 gttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaag 2601
          || || |||  |||   ||| ||||||||| ||  |  | | ||| ||| ||||||||||
Db   5816 GTCGGGGATCAAGTGCTTGTCAGACGCCATCGACCCAGCAGCCTTGAGCCTCGGTGGAAA 5875

Qy   2602 ggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatccccacc 2661
          || || || || ||  |  |||| || || || || || ||||||||| || ||  |  ||
Db   5876 GGCCCATACCTGGTGTTGCTGACTACCCCGACCGCGGTAAAAGTCGATGGTATTGCTGCC 5935

Qy   2662 tggatccatgcatcccacgttaagccggcgccacct------cccgattcggggtggaaa 2715
          ||| ||||||| || ||| | || || || ||||||      || |||  |   ||| |
Db   5936 TGGGTCCATGCTTCTCACCTCAAACCTGCACCACCTTCGGCACCAGATGAGTCCTGGGAG 5995

Qy   2716 gccgaaaagactgaaaatccccttaagcttcgcctccatcgcgtggttccttactctgtc 2775
           ||||||||||  |||| |||||||| ||  | |  ||  |      | ||||
Db   5996 CTGGAAAAGACTGATCATCCTCTTAAGCTGCGTATTCGGCGGCGGCGGGACGAGTCTGCA 6055

Qy   2776 aa 2777
          ||
Db   6056 AA 6057
```

```
RESULT   7
US-08-110-300A-8
; Sequence 8, Application US/08110300A
; Patent No. 5643756
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,300A
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
```

```
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-110-300A-8


  Query Match             28.1%;  Score 933.6;  DB 1;  Length 8323;
  Best Local Similarity   60.0%;  Pred. No. 1.2e-273;
  Matches 1702;  Conservative    0;  Mismatches 1094;  Indels   42;  Gaps    7;

Qy     19 acagatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcgaa 78
          | |||||| || |||||||||||| | |||||||| || ||  | || || ||||| ||
Db   3049 AAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCAGTCAGTCTCTCTTCGCCTTTGAG 3108

Qy     79 tggagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaaggg 138
          ||||||||||||  | |||||   | || ||  | |||||||| |||| || || ||
Db   3109 TGGAGAGATCCAGAGATGGGAATCTCAGGACAATTAACCTGGACCAGACTCCCGCAGGGT 3168

Qy    139 ttcaagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcagg 198
          ||||| |||   || ||| | ||||| |||||||| |||||||||| ||  ||||| ||
Db   3169 TTCAAAAACAGTCCCACCCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGG 3228

Qy    199 atccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagcc 258
          ||||| |||||  |  |||  || |||||||| || |||||| | || |||||  |   |
Db   3229 ATCCAGCACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCCACT 3288

Qy    259 accaaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggc 318
          |  | |  |||||   | ||||||||  ||| ||  | |       |    ||||| ||
Db   3289 TCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCTAGGGGACCTCGGA 3348

Qy    319 tacagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtac 378
          ||  | ||||| || ||||| ||||| ||||||  || | |||| |  ||  |||||||
Db   3349 TATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTAT 3408

Qy    379 agtttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagata 438
            | |    |  || ||| |||||||||| |||||  | ||  | |||||  |   |
Db   3409 CTTCTAAAAGAGGGTCAGAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAG 3468

Qy    439 ccggccccaaccacagccaaacaagtgagagagtttttggggacagctggatttgcaga 498
          ||  | || |  ||  |    |||| | || |||||  | ||||| || || || ||  |
Db   3469 CCTACTCCGAAGACCCCTCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGC 3528

Qy    499 ctgtggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaa 558
          || |||||||| ||||||||      | ||||||||  | ||||| || ||||||
Db   3529 CTCTGGATCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACGGGG 3588

Qy    559 gggggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctg 618
                ||    ||||  || || || || ||||| |   | |  |||||  |||| ||
Db   3589 ACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGATCAAGCAGGCTCTC 3648

Qy    619 ctgagcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggat 678
            | |  || ||||| ||||   | || ||| | ||||| |||||    |||| ||| ||
Db   3649 TTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTAAGCCCTTCGAACTTTTTGTTGAC 3708

Qy    679 gagcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacct 738
          |||    ||||    |||  ||| ||  |||| ||||  || || || ||| |  | ||
Db   3709 GAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAAACTGGGGCCTTGGCGTCGGCCG 3768

Qy    739 gttgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaag 798
          || |||||||||| || ||||| || || || ||   ||| ||||||   || ||   |
Db   3769 GTGGCCTACCTGTCCAAAAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGG 3828

Qy    799 gctatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaat 858
             | |||||  | ||| | |||  ||| ||||||| |||  | ||  ||||||||
Db   3829 ATGGTAGCAGCCATCGCCGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAGCCA 3888

Qy    859 ataactgtaatagccccccatgcattggagaacatcgttcggcagcccccagaccgatgg 918
```



RESULT 8
US-08-886-642-8
; Sequence 8, Application US/08886642
; Patent No. 5952474
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:

```
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  45 Rockefeller Plaza, Suite 2800
;       CITY:  New York
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10111
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/886,642
;       FILING DATE:  01-JUL-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/110,300
;       FILING DATE:  20-AUG-1993
;       APPLICATION NUMBER:  07/938,100
;       FILING DATE:  28-AUG-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hone, William J.
;       REGISTRATION NUMBER:  26,739
;       REFERENCE/DOCKET NUMBER:  07763/010002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212/765-5070
;       TELEFAX:  212/258-2291
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8323 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
US-08-886-642-8
```

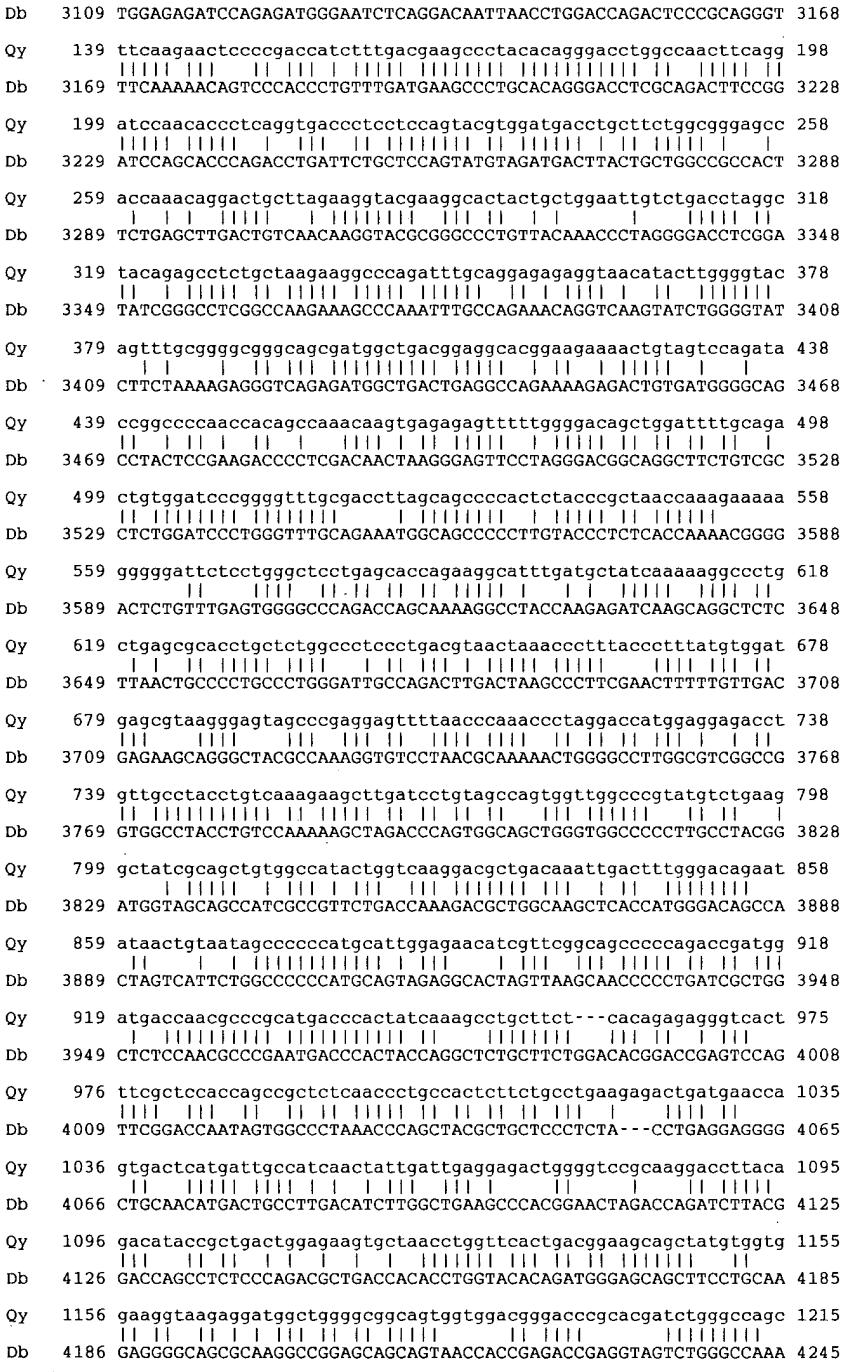
```
  Query Match             28.1%;  Score 933.6;  DB 2;  Length 8323;
  Best Local Similarity   60.0%;  Pred. No. 1.2e-273;
  Matches 1702;  Conservative    0;  Mismatches 1094;  Indels   42;  Gaps    7;

Qy      19 acagatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaa 78
           | |||||| || |||||||||||| | |||||||| || ||  | || || ||||| ||
Db    3049 AAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCAGTCAGTCTCTCTTCGCCTTTGAG 3108

Qy      79 tggagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaaggg 138
           ||||||||||||||  | |||||   | || ||  | |||||||||| |||| || || ||
Db    3109 TGGAGAGATCCAGAGATGGGAATCTCAGGACAATTAACCTGGACCAGACTCCCGCAGGGT 3168

Qy     139 ttcaagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcagg 198
           ||||| |||   || ||| | ||||| |||||||| |||||||||| ||  ||||| ||
Db    3169 TTCAAAAACAGTCCCACCCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGG 3228

Qy     199 atccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagcc 258
           ||||| |||||  |  |||  || |||||||| || ||||||  | || |||||  |   |
Db    3229 ATCCAGCACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCCACT 3288

Qy     259 accaaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggc 318
           |  | |  |||||   | ||||||||  ||| ||  | |      |    ||||| ||
Db    3289 TCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCTAGGGGACCTCGGA 3348

Qy     319 tacagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtac 378
           ||  | ||||| || ||||| ||||| ||||||  || | |||| |  ||  |||||||
Db    3349 TATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTAT 3408

Qy     379 agtttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagata 438
           | | |  | || ||| |||||||||| |||||  | ||  | |||||  |   |
Db    3409 CTTCTAAAAGAGGGTCAGAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAG 3468

Qy     439 ccggccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcaga 498
           ||  | || |  ||  |   |||| | || |||||  | ||||| || || || ||  |
Db    3469 CCTACTCCGAAGACCCCTCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGC 3528

Qy     499 ctgtggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaa 558
```

```
           || |||||||| |||||||| | |||||||| | ||||| || ||||||
Db    3529 CTCTGGATCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACGGGG 3588

Qy     559 gggggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctg 618
                 ||    |||| || || || || ||||| |   | |  |||||  |||| ||
Db    3589 ACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGATCAAGCAGGCTCTC 3648

Qy     619 ctgagcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggat 678
           | |  || ||||| |||| | || ||| | ||||| |||||    |||| ||| ||
Db    3649 TTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTAAGCCCTTCGAACTTTTTGTTGAC 3708

Qy     679 gagcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacct 738
           |||    ||||    |||  ||| ||  |||| |||| || || || ||| |  | ||
Db    3709 GAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAAACTGGGGCCTTGGCGTCGGCCG 3768

Qy     739 gttgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaag 798
           || |||||||||| || ||||| || || || ||  ||| ||||||   || ||   |
Db    3769 GTGGCCTACCTGTCCAAAAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGG 3828

Qy     799 gctatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaat 858
           |  |||||  | ||| | |||  ||| ||||||| |||  | ||  ||||||||
Db    3829 ATGGTAGCAGCCATCGCCGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAGCCA 3888

Qy     859 ataactgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatgg 918
           ||    |  | |||||||||||| | |||    | |||  ||| ||||| || || |||
Db    3889 CTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTTAAGCAACCCCCTGATCGCTGG 3948

Qy     919 atgaccaacgcccgcatgacccactatcaaagcctgcttct---cacagagagggtcact 975
           |  |||||||||| |||||||||| ||    ||||||||   ||| ||  | |||
Db    3949 CTCTCCAACGCCCGAATGACCCACTACCAGGCTCTGCTTCTGGACACGGACCGAGTCCAG 4008

Qy     976 ttcgctccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaacca 1035
           ||||  |||  ||  || || ||||| || || || || |||  |   |||| ||
Db    4009 TTCGGACCAATAGTGGCCCTAAACCCAGCTACGCTGCTCCCTCTA---CCTGAGGAGGGG 4065

Qy    1036 gtgactcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttaca 1095
           ||   ||||| |||| | |  | |||  |||  |     ||     |    || |||||
Db    4066 CTGCAACATGACTGCCTTGACATCTTGGCTGAAGCCCACGGAACTAGACCAGATCTTACG 4125

Qy    1096 gacataccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtg 1155
           |||   || ||  | |  |  |  | ||||||| ||| || || |||||||   ||
Db    4126 GACCAGCCTCTCCCAGACGCTGACCACACCTGGTACACAGATGGGAGCAGCTTCCTGCAA 4185

Qy    1156 gaaggtaagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagc 1215
           || ||  || | | ||| || || |||||      || ||||       |||||||||
Db    4186 GAGGGGCAGCGCAAGGCCGGAGCAGCAGTAACCACCGAGACCGAGGTAGTCTGGGCCAAA 4245

Qy    1216 agcctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttg 1275
           |||||  |  || || || || ||||  ||||||  | || || |||||  |||| ||
Db    4246 GCACTGCCAGCCGGGACATCGGCCCAAAGAGCTGAGTTGATAGCGCTCACCCAAGCCTTA 4305

Qy    1276 cggctggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgact 1335
              ||||  |||||  ||     | ||  |||| || || ||| | ||||| ||||| |||
Db    4306 AAAATGGCAGAAGGTAAGAAGCTGAATGTTTACACCGATAGCCGTTATGCTTTTGCCACT 4365

Qy    1336 gcacacgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaa 1395
           || ||  | ||||| |  || |||| |     | |||||||| || |||| ||| |  |||
Db    4366 GCCCATATTCACGGAGAAATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAA 4425

Qy    1396 ataaagaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggcta 1455
           || || || || || ||  |   |||| |  || |     |||| ||||| ||
Db    4426 ATCAAAAATAAGGACGAGATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGACTT 4485

Qy    1456 gctattatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccag 1515
              || || || || ||||||||||| |  ||          | || || |||  |
Db    4486 AGCATAATTCATTGCCCGGGACATCAGAAGGGAAACCGCGCGGAGGCAAGGGGCAACAGG 4545

Qy    1516 atggctgaccgggttgccaagcaggcagcccaggctgttaacccttctgcctataatagaa 1575
           |||||  ||||  |  |||    | |  ||||       |  |  |    | ||  | |
Db    4546 ATGGCCGACCAAGCGGCCCGAGAAGTAGCCACTAGAGAAACTCCAGAGACTTCCACACTT 4605

Qy    1576 acgcccaaagccccagaaccc-agacgacagtacaccctagaagactggcaagagataaa 1634
              |    ||    |||  ||| | ||    | |||   |        ||  |  || | |
```

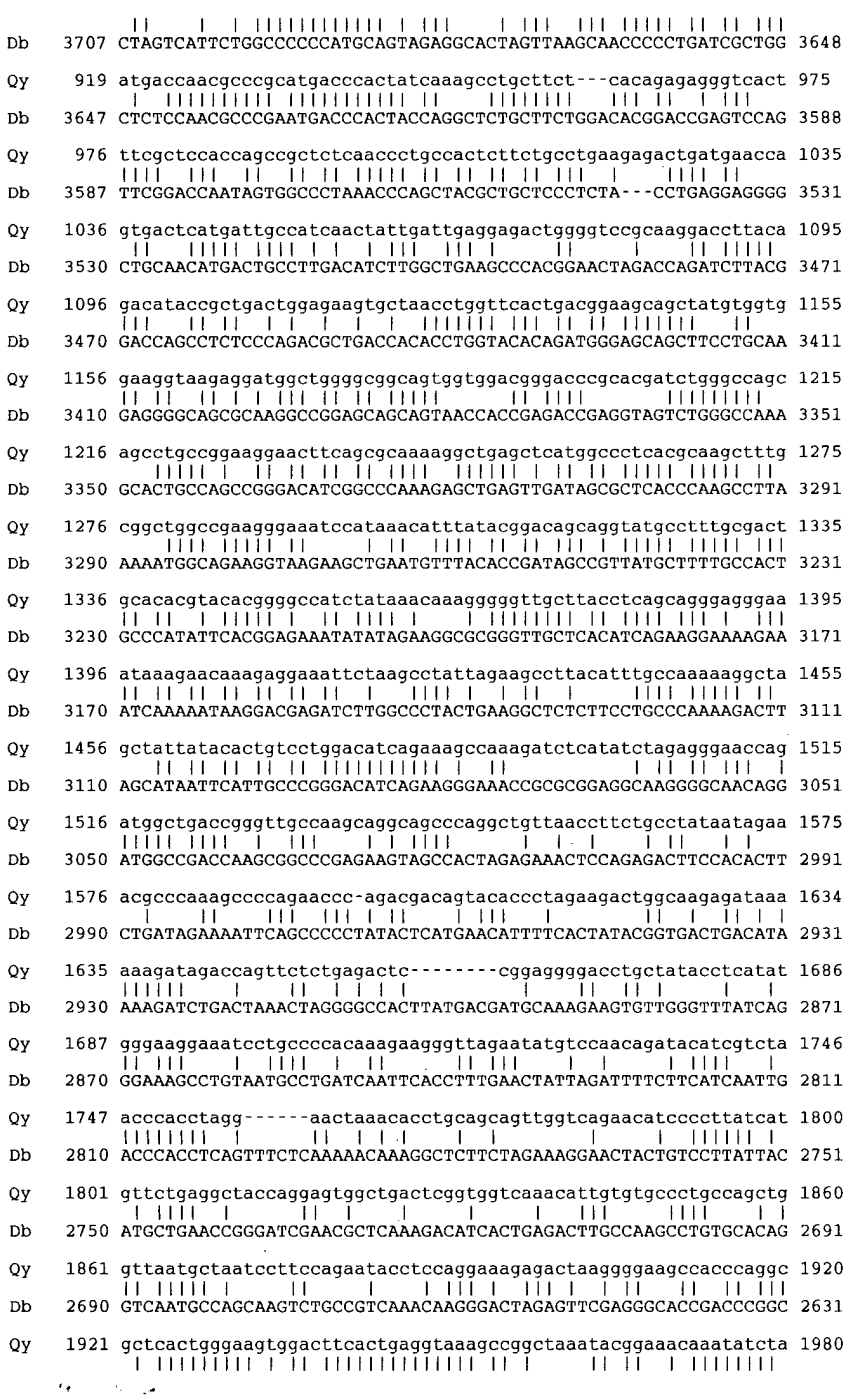
```
Db   4606 CTGATAGAAAATTCAGCCCCCTATACTCATGAACATTTTCACTATACGGTGACTGACATA 4665

Qy   1635 aaagatagaccagttctctgagactc--------cggaggggacctgctatacctcatat 1686
          ||||||    |   ||  | | | |        |   ||  || |     |   |
Db   4666 AAAGATCTGACTAAACTAGGGGCCACTTATGACGATGCAAAGAAGTGTTGGGTTTATCAG 4725

Qy   1687 gggaaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatcgtcta 1746
          || |||    |  |||| |  ||      || |||    |  |     | ||||   |
Db   4726 GGAAAGCCTGTAATGCCTGATCAATTCACCTTTGAACTATTAGATTTTCTTCATCAATTG 4785

Qy   1747 acccacctagg------aactaaacacctgcagcagttggtcagaacatccccttatcat 1800
          ||||||||  |      ||  | | |    |  |        |     |  |||||| |
Db   4786 ACCCACCTCAGTTTCTCAAAAACAAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTAC 4845

Qy   1801 gttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctg 1860
          | ||||  |      ||  |   |     |     |   |||     ||||    | |
Db   4846 ATGCTGAACCGGGATCGAACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAG 4905

Qy   1861 gttaatgctaatccttccagaatacctccaggaaagagactaaggggaagccacccaggc 1920
          || ||||| |     ||      |    | ||| |  ||| |  | ||   ||  || |||
Db   4906 GTCAATGCCAGCAAGTCTGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCGACCCGGC 4965

Qy   1921 gctcactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatcta 1980
          | |||||||| | || ||||||||||||||| || |     || ||  | ||||||||
Db   4966 ACCCACTGGGAAATTGATTTCACTGAGGTAAAACCTGGCCTGTATGGGTATAAATATCTT 5025

Qy   1981 ttggtttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagact 2040
          || |||||  ||||||| || || |||||||||||| ||||  || || |||||||| |||
Db   5026 TTAGTTTTCATAGACACTTTCTCTGGATGGGTAGAAGCTTTCCCAACCAAGAAAGAAACT 5085

Qy   2041 tcaaccgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaag 2100
          | |  || ||  | |||||  |||| || ||||| || || ||||| || || ||  ||
Db   5086 GCCAAAGTTGTAACCAAGAAGCTACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAG 5145

Qy   2101 gtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaag 2160
          ||| | ||  | ||||||| || || |||||  || ||||||||||  | | ||| |
Db   5146 GTATTGGGAACCGACAATGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGAT 5205

Qy   2161 atattggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggta 2220
          || ||||| |||||||||||| |||||||| |||||||||| || ||||| ||||||
Db   5206 TTACTGGGGGTTGATTGGAAACTACATTGTGCTTACAGACCCCAGAGTTCAGGTCAGGTA 5265

Qy   2221 gagaggatgaatagaaccattaaagagacccttaccaaattgaccacagagactggcatt 2280
          || || |||||||| || || || |||||  | || ||||||||    |  ||||||  |
Db   5266 GAAAGAATGAATAGGACAATCAAGGAGACTTTAACTAAATTGACGCTTGCAACTGGCTCT 5325

Qy   2281 aatgattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacag 2340
          |  || ||| ||   || || ||| | |  || | | |   | |||||  || || |
Db   5326 AGGGACTGGGTGCTCCTGCTTCCCCTAGCCCTGTATCGAGCCCGCAACACGCCGGGCCCC 5385

Qy   2341 tttgggctgacccccctatgaattgctctacgggggaccccccccgttggcagaaattgcc 2400
          |||  || ||||| ||||||  |  | || |||| |||||| ||  | | | |  | |
Db   5386 CATGGTCTCACCCCATATGAAATCTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCT 5445

Qy   2401 tttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctc 2460
          | |  | |  ||  | ||    | |  ||   || |      |    |  |||| |||
Db   5446 GATCCTGACATGGCAAAGGTTACTCATAACCCCTCTCTCCAAGCCCATTTACAGGCACTC 5505

Qy   2461 gagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagacttg 2520
          |   |||   |||   ||  ||||  | | |   ||  || |||  || |  |
Db   5506 TACCTGGTCCAGCACGAAGTCTGGAGACCGTTGGCGGCAGCTTACCAAGAACAACTGGAC 5565

Qy   2521 c------aagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgt 2574
          |       ||| || || |  |||| ||| || ||  |||| |  ||  |  | ||||
Db   5566 CGGCCGGTAGTGCCTCACCCTTTCCGAGTCGGTGACACAGTGTGGGTCCGCAGACACCAA 5625

Qy   2575 gcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacg 2634
          |   ||| || ||  | || ||||| |||||||||  ||| ||  |||| || || ||
Db   5626 ACTAAAAATCTAGAACCCCGCTGGAAAGGACCTTATACCGTCCTACTGACTACCCCCACC 5685

Qy   2635 gctgtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccggc---- 2690
          ||| | ||||| || || ||  |  | |||||||| ||  ||||||| ||| | ||
Db   5686 GCTCTCAAAGTGGACGGCATTGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGAC 5745

Qy   2691 ----------gccacctcccgattcggggtggaaagccgaaaagactgaaaatccccttt 2739
                     ||||||  | || |||   |||   |   ||    || ||||||||||
Db   5746 ACCAGGATTGAGCCACCATCGGAATCGACATGGCGTGTTCAACGCTCTCAAAATCCCCTA 5805

Qy   2740 aagcttcgcctccatcgcgtggttccttactctgtcaataactcctcaagttaatggtaa 2799
          |||  |  |  |    |||| |    | || ||      |   ||| ||  ||   |  |
Db   5806 AAGATAAGATTGACCCGCGGGACCTCCTAATCCCCTTAATTCTCTTCCTGTCTCTCAAAG 5865

Qy   2800 acgccttgtggacagccc 2817
             |||   |   |||| |
Db   5866 GGGCCAGATCCGCAGCAC 5883
```

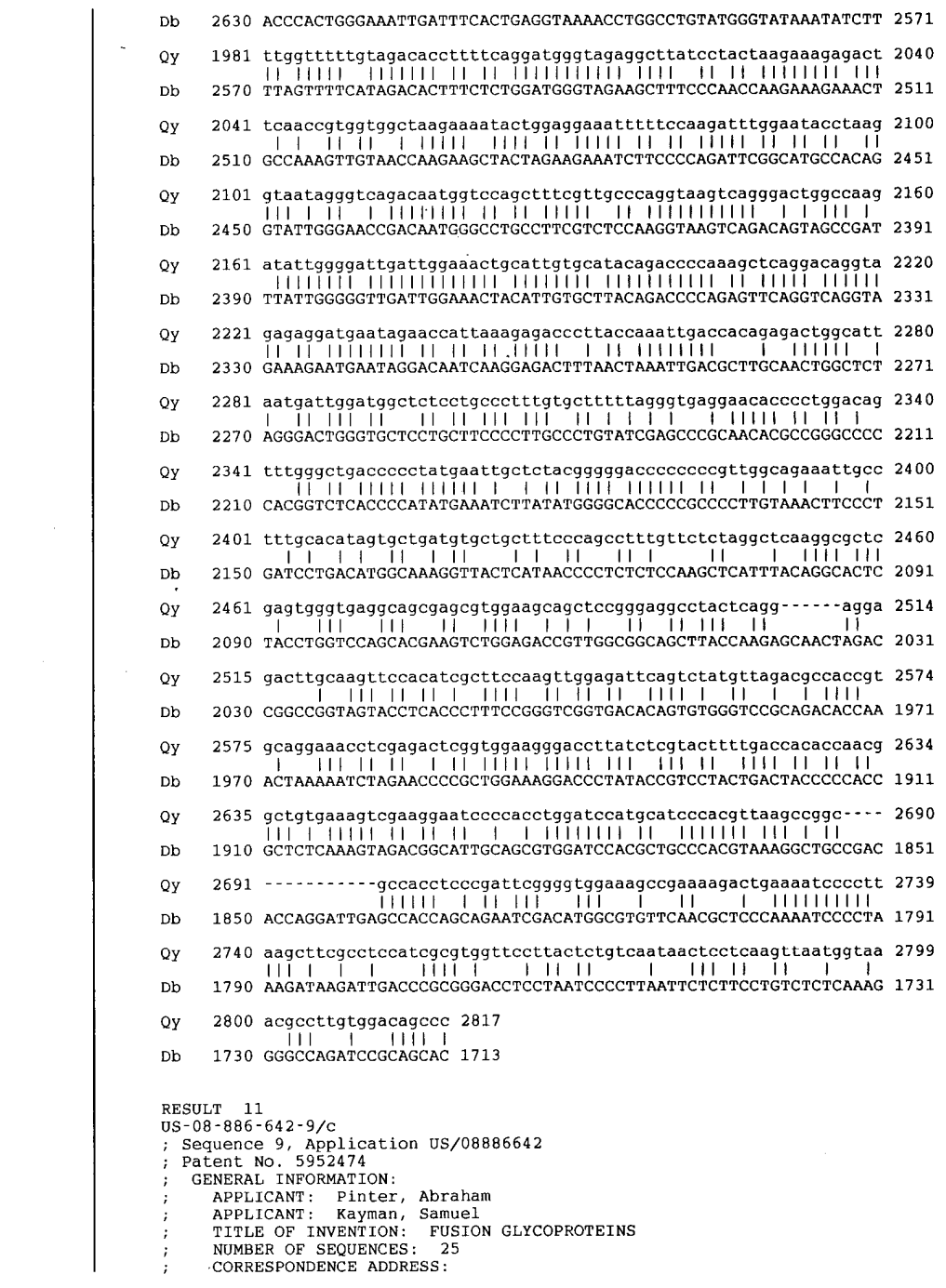
```
RESULT   9
PCT-US93-08041-8
; Sequence 8, Application PC/TUS9308041
;  GENERAL INFORMATION:
;    APPLICANT:  The Public Health Research Institute of the City of
;    APPLICANT:  New York, Inc.
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08041
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/938,100
;      FILING DATE:  28-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/_____
;      FILING DATE:  20-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-08041-8
```

```
  Query Match             28.1%;  Score 933.6;  DB 5;  Length 8323;
  Best Local Similarity   60.0%;  Pred. No. 1.2e-273;
  Matches 1702;  Conservative    0;  Mismatches 1094;  Indels   42;  Gaps    7;

Qy      19 acagatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaa 78
           | |||||| || |||||||||||| | |||||||| || ||  | || || |||| ||
Db    3049 AAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCAGTCAGTCTCTCTTCGCCTTTGAG 3108

Qy      79 tggagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaaggg 138
           ||||||||||||||  | |||||   | || ||  | |||||||||| |||| || || ||
```


Db 3109 TGGAGAGATCCAGAGATGGGAATCTCAGGACAATTAACCTGGACCAGACTCCCGCAGGGT 3168
Qy 139 ttcaagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcagg 198
Db 3169 TTCAAAAACAGTCCCACCCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGG 3228
Qy 199 atccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagcc 258
Db 3229 ATCCAGCACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCCACT 3288
Qy 259 accaaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggc 318
Db 3289 TCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCTAGGGGACCTCGGA 3348
Qy 319 tacagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtac 378
Db 3349 TATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTAT 3408
Qy 379 agtttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagata 438
Db 3409 CTTCTAAAAGAGGGTCAGAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAG 3468
Qy 439 ccggccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcaga 498
Db 3469 CCTACTCCGAAGACCCCTCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGC 3528
Qy 499 ctgtggatcccgggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaa 558
Db 3529 CTCTGGATCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACGGGG 3588
Qy 559 gggggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctg 618
Db 3589 ACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGATCAAGCAGGCTCTC 3648
Qy 619 ctgagcgcacctgctctggccctccctgacgtaactaaaccctttacccttatgtggat 678
Db 3649 TTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTAAGCCCTTCGAACTTTTTGTTGAC 3708
Qy 679 gagcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacct 738
Db 3709 GAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAAACTGGGGCCTTGGCGTCGGCCG 3768
Qy 739 gttgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaag 798
Db 3769 GTGGCCTACCTGTCCAAAAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGG 3828
Qy 799 gctatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaat 858
Db 3829 ATGGTAGCAGCCATCGCCGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAGCCA 3888
Qy 859 ataactgtaatagccccccatgcattggagaacatcgttcggcagcccccagaccgatgg 918
Db 3889 CTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTTAAGCAACCCCCTGATCGCTGG 3948
Qy 919 atgaccaacgcccgcatgacccactatcaaagcctgcttct---cacagagagggtcact 975
Db 3949 CTCTCCAACGCCCGAATGACCCACTACCAGGCTCTGCTTCTGGACACGGACCGAGTCCAG 4008
Qy 976 ttcgctccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaacca 1035
Db 4009 TTCGGACCAATAGTGGCCCTAAACCCAGCTACGCTGCTCCCTCTA---CCTGAGGAGGGG 4065
Qy 1036 gtgactcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttaca 1095
Db 4066 CTGCAACATGACTGCCTTGACATCTTGGCTGAAGCCCACGGAACTAGACCAGATCTTACG 4125
Qy 1096 gacataccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtg 1155
Db 4126 GACCAGCCTCTCCCAGACGCTGACCACACCTGGTACACAGATGGGAGCAGCTTCCTGCAA 4185
Qy 1156 gaaggtaagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagc 1215
Db 4186 GAGGGGCAGCGCAAGGCCGGAGCAGCAGTAACCACCGAGACCGAGGTAGTCTGGGCCAAA 4245
Qy 1216 agcctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttg 1275
Db 4246 GCACTGCCAGCCGGGACATCGGCCCAAAGAGCTGAGTTGATAGCGCTCACCCAAGCCTTA 4305
Qy 1276 cggctggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgact 1335
Db 4306 AAAATGGCAGAAGGTAAGAAGCTGAATGTTTACACCGATAGCCGTTATGCTTTTGCCACT 4365
Qy 1336 gcacacgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaa 1395
Db 4366 GCCCATATTCACGGAGAAATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAA 4425
Qy 1396 ataaagaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggcta 1455
Db 4426 ATCAAAAATAAGGACGAGATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGACTT 4485
Qy 1456 gctattatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccag 1515
Db 4486 AGCATAATTCATTGCCCGGGACATCAGAAGGGAAACCGCGCGGAGGCAAGGGGCAACAGG 4545
Qy 1516 atggctgaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaa 1575
Db 4546 ATGGCCGACCAAGCGGCCCGAGAAGTAGCCACTAGAGAAACTCCAGAGACTTCCACACTT 4605
Qy 1576 acgcccaaagccccagaaccc-agacgacagtacaccctagaagactggcaagagataaa 1634
Db 4606 CTGATAGAAAATTCAGCCCCCTATACTCATGAACATTTTCACTATACGGTGACTGACATA 4665
Qy 1635 aaagatagaccagttctctgagactc--------cggaggggacctgctatacctcatat 1686
Db 4666 AAAGATCTGACTAAACTAGGGGCCACTTATGACGATGCAAAGAAGTGTTGGGTTTATCAG 4725
Qy 1687 gggaaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatcgtcta 1746
Db 4726 GGAAAGCCTGTAATGCCTGATCAATTCACCTTTGAACTATTAGATTTTCTTCATCAATTG 4785
Qy 1747 acccacctagg------aactaaacacctgcagcagttggtcagaacatccccttatcat 1800
Db 4786 ACCCACCTCAGTTTCTCAAAAACAAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTAC 4845
Qy 1801 gttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctg 1860
Db 4846 ATGCTGAACCGGGATCGAACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAG 4905
Qy 1861 gttaatgctaatccttccagaatacctccaggaaagagactaaggggaagccacccaggc 1920
Db 4906 GTCAATGCCAGCAAGTCTGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCGACCCGGC 4965
Qy 1921 gctcactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatcta 1980
Db 4966 ACCCACTGGGAAATTGATTTCACTGAGGTAAAACCTGGCCTGTATGGGTATAAATATCTT 5025
Qy 1981 ttggtttttgtagacacctttcaggatgggtagaggcttatcctactaagaaagagact 2040
Db 5026 TTAGTTTTCATAGACACTTTCTCTGGATGGGTAGAAGCTTTCCCAACCAAGAAAGAAACT 5085
Qy 2041 tcaaccgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaag 2100
Db 5086 GCCAAAGTTGTAACCAAGAAGCTACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAG 5145
Qy 2101 gtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaag 2160
Db 5146 GTATTGGGAACCGACAATGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGAT 5205
Qy 2161 atattggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggta 2220
Db 5206 TTACTGGGGGTTGATTGGAAACTACATTGTGCTTACAGACCCCAGAGTTCAGGTCAGGTA 5265
Qy 2221 gagaggatgaatagaaccattaaagagacccttaccaaattgaccacagagactggcatt 2280
Db 5266 GAAAGAATGAATAGGACAATCAAGGAGACTTTAACTAAATTGACGCTTGCAACTGGCTCT 5325

```
Qy   2281 aatgattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacag 2340
          |  || ||| ||   || || ||| | |  || | | | |   | ||||| || || |
Db   5326 AGGGACTGGGTGCTCCTGCTTCCCCTAGCCCTGTATCGAGCCCGCAACACGCCGGGCCCC 5385

Qy   2341 tttgggctgacccccctatgaattgctctacgggggaccccccccgttggcagaaattgcc 2400
            ||| || ||||| |||||| |  | || |||| |||||| ||  | | | |  |  |
Db   5386 CATGGTCTCACCCCATATGAAATCTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCT 5445

Qy   2401 tttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctc 2460
            | |  | |  ||  | ||    | |  ||  || |      |    |  |||| |||
Db   5446 GATCCTGACATGGCAAAGGTTACTCATAACCCCTCTCTCCAAGCCCATTTACAGGCACTC 5505

Qy   2461 gagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagacttg 2520
            |   |||   |||   ||  ||||  | | |   ||  || |||  || |  |
Db   5506 TACCTGGTCCAGCACGAAGTCTGGAGACCGTTGGCGGCAGCTTACCAAGAACAACTGGAC 5565

Qy   2521 c------aagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgt 2574
          |       ||| || || |  |||| ||| || ||  |||| |  ||  |  | ||||
Db   5566 CGGCCGGTAGTGCCTCACCCTTTCCGAGTCGGTGACACAGTGTGGGTCCGCAGACACCAA 5625

Qy   2575 gcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacg 2634
            |   ||| || ||  | || ||||| |||||||||  ||| ||  |||| || || ||
Db   5626 ACTAAAAATCTAGAACCCCGCTGGAAAGGACCTTATACCGTCCTACTGACTACCCCCACC 5685

Qy   2635 gctgtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccggc---- 2690
          ||| | ||||| || || ||  |  | ||||||| ||  ||||||| ||| | ||
Db   5686 GCTCTCAAAGTGGACGGCATTGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGAC 5745

Qy   2691 ----------gccacctcccgattcggggtggaaagccgaaaagactgaaaatccccttt 2739
                     ||||||  | || |||   |||   |   ||    || ||||||||||
Db   5746 ACCAGGATTGAGCCACCATCGGAATCGACATGGCGTGTTCAACGCTCTCAAAATCCCCTA 5805

Qy   2740 aagcttcgcctccatcgcgtggttccttactctgtcaataactcctcaagttaatggtaa 2799
          ||| |  |  |    |||| |    | || ||      |   ||| ||  ||   |  |
Db   5806 AAGATAAGATTGACCCGCGGGACCTCCTAATCCCCTTAATTCTCTTCCTGTCTCTCAAAG 5865

Qy   2800 acgccttgtggacagccc 2817
            |||   |  |||| |
Db   5866 GGGCCAGATCCGCAGCAC 5883


RESULT  10
US-08-110-300A-9/c
; Sequence 9, Application US/08110300A
; Patent No. 5643756
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,300A
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-110-300A-9


  Query Match             28.1%;  Score 932;  DB 1;  Length 10367;
  Best Local Similarity   59.9%;  Pred. No. 4.3e-273;
  Matches 1701;  Conservative    0;  Mismatches 1095;  Indels   42;  Gaps    7;

Qy     19 acagatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaa 78
          | |||||| || |||||||||||| | |||||||| || ||  | || || ||||| ||
Db   4547 AAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCAGTCAGTCTCTCTTCGCCTTTGAG 4488

Qy     79 tggagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaaggg 138
          ||||||||||||||  | |||||   | || ||  | ||||||||| |||| || || ||
Db   4487 TGGAGAGATCCAGAGATGGGAATCTCAGGACAATTAACCTGGACCAGACTCCCGCAGGGT 4428

Qy    139 ttcaagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcagg 198
          ||||| |||   || ||| | ||||| |||||||| |||||||||| ||  ||||| ||
Db   4427 TTCAAAAACAGTCCCACCCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGG 4368

Qy    199 atccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagcc 258
          ||||| |||||  |  |||  || |||||||| || ||||||  | || |||||  |   |
Db   4367 ATCCAGCACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCCACT 4308

Qy    259 accaaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggc 318
            | | |  |||||   | ||||||||  ||| ||  | |      |    ||||| ||
Db   4307 TCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCTAGGGGACCTCGGA 4248

Qy    319 tacagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtac 378
          ||  | ||||| || ||||| ||||| ||||||  || | |||| |  ||  |||||||
Db   4247 TATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTAT 4188

Qy    379 agtttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagata 438
            | |    |  || ||| |||||||||| |||||  | ||  | |||||  |   |
Db   4187 CTTCTAAAAGAGGGTCAGAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAG 4128

Qy    439 ccggccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcaga 498
          ||  | || |  ||  |   |||| | || |||||  | |||||  || || || ||  |
Db   4127 CCTACTCCGAAGACCCCTCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGC 4068

Qy    499 ctgtggatcccgggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaa 558
          || |||||||| ||||||||      | ||||||||  | ||||| || |||||||
Db   4067 CTCTGGATCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACGGGG 4008

Qy    559 gggggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctg 618
                ||    ||||  || || || || ||||| |   | |  |||||  |||| ||
Db   4007 ACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGATCAAGCAGGCTCTC 3948

Qy    619 ctgagcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggat 678
            | |  || ||||| ||||   | || ||| | ||||| |||||    |||| ||| ||
Db   3947 TTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTAAGCCCTTCGAACTTTTTGTTGAC 3888

Qy    679 gagcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacct 738
          |||    ||||    |||  ||| ||  |||| ||||  || || || ||| |  | ||
Db   3887 GAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAAACTGGGGCCTTGGCGTCGGCCG 3828

Qy    739 gttgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaag 798
          || |||||||||| || ||||| || || || ||   ||| ||||||   || ||   |
Db   3827 GTGGCCTACCTGTCCAAAAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGG 3768

Qy    799 gctatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaat 858
            | ||||| | ||| | |||  ||| ||||||| |||  | ||  ||||||||
Db   3767 ATGGTAGCAGCCATCGCCGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAGCCA 3708

Qy    859 ataactgtaatagcccccatgcattggagaacatcgttcggcagcccccagaccgatgg 918
```



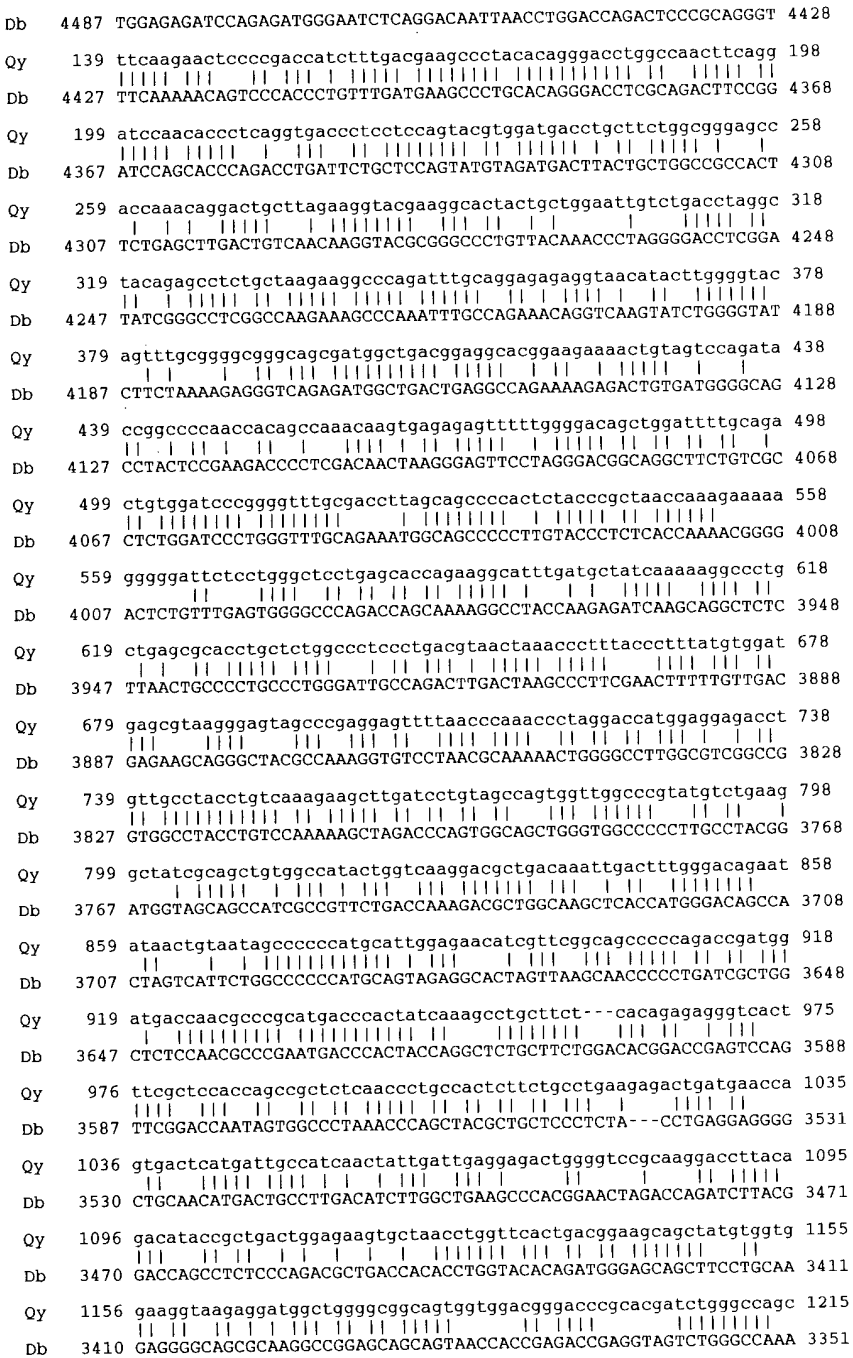


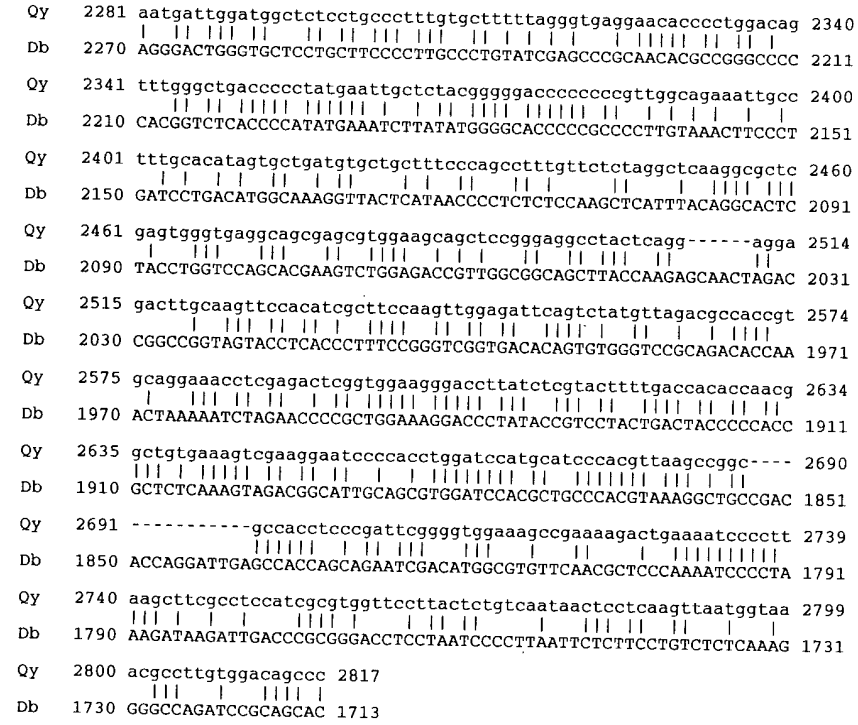
RESULT 11
US-08-886-642-9/c
; Sequence 9, Application US/08886642
; Patent No. 5952474
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:

```
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  45 Rockefeller Plaza, Suite 2800
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/886,642
;      FILING DATE:  01-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/110,300
;      FILING DATE:  20-AUG-1993
;      APPLICATION NUMBER:  07/938,100
;      FILING DATE:  28-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hone, William J.
;      REGISTRATION NUMBER:  26,739
;      REFERENCE/DOCKET NUMBER:  07763/010002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212/765-5070
;      TELEFAX:  212/258-2291
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-886-642-9


  Query Match             28.1%;  Score 932;  DB 2;  Length 10367;
  Best Local Similarity   59.9%;  Pred. No. 4.3e-273;
  Matches 1701;  Conservative    0;  Mismatches 1095;  Indels   42;  Gaps    7;

Qy      19 acagatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcgaa 78
           | |||||| || ||||||||||||| | |||||||| || ||  | || || ||||| ||
Db    4547 AAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCAGTCAGTCTCTCTTCGCCTTTGAG 4488

Qy      79 tggagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaaggg 138
           |||||||||||||  | |||||   | || ||  | ||||||||| |||| || || ||
Db    4487 TGGAGAGATCCAGAGATGGGAATCTCAGGACAATTAACCTGGACCAGACTCCCGCAGGGT 4428

Qy     139 ttcaagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcagg 198
           ||||| |||   || ||| | ||||| |||||||| |||||||||| ||  ||||| ||
Db    4427 TTCAAAAACAGTCCCACCCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGG 4368

Qy     199 atccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagcc 258
           ||||| ||||| | ||| || |||||||| || |||||| | || ||||| |   |
Db    4367 ATCCAGCACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCCACT 4308

Qy     259 accaaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggc 318
           |  | |  |||||   | ||||||||  ||| ||  | |      |   |||||| ||
Db    4307 TCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCTAGGGGACCTCGGA 4248

Qy     319 tacagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtac 378
           ||  | ||||| || ||||| ||||| ||||||  || | |||| |  ||  |||||||
Db    4247 TATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTAT 4188

Qy     379 agtttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagata 438
           | |   |  || ||| |||||||||| |||||  | ||  | |||||  |   |
Db    4187 CTTCTAAAAGAGGGTCAGAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAG 4128

Qy     439 ccggccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcaga 498
           ||  | || |  ||  |   |||| | || |||||  | ||||| || || || ||  |
Db    4127 CCTACTCCGAAGACCCCTCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGC 4068

Qy     499 ctgtggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaa 558
           || |||||||| ||||||||     | ||||||||  | |||||| || ||||||
Db    4067 CTCTGGATCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACGGGG 4008

Qy     559 gggggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctg 618
           ||    ||||  || || || || ||||| |   | |  |||||  |||| ||
Db    4007 ACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGATCAAGCAGGCTCTC 3948

Qy     619 ctgagcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggat 678
           | |  || ||||| |||    | || ||| | ||||| |||||     |||| ||| ||
Db    3947 TTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTAAGCCCTTCGAACTTTTTGTTGAC 3888

Qy     679 gagcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacct 738
           |||    ||||    |||  ||| ||  |||| ||||  || || || ||| |  | ||
Db    3887 GAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAAACTGGGGCCTTGGCGTCGGCCG 3828

Qy     739 gttgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaag 798
           || |||||||||||  || ||||| || || || ||   ||| ||||||   || ||   |
Db    3827 GTGGCCTACCTGTCCAAAAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGG 3768

Qy     799 gctatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaat 858
           | |||||  | ||| | |||  ||| ||||||| |||  | ||  ||||||||
Db    3767 ATGGTAGCAGCCATCGCCGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAGCCA 3708

Qy     859 ataactgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatgg 918
           ||    |  | |||||||||||| | |||    | |||  ||| ||||| || || |||
Db    3707 CTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTTAAGCAACCCCCTGATCGCTGG 3648

Qy     919 atgaccaacgcccgcatgacccactatcaaagcctgcttct---cacagagagggtcact 975
           |  |||||||||| |||||||||| ||    ||||||||    ||| ||  | |||
Db    3647 CTCTCCAACGCCCGAATGACCCACTACCAGGCTCTGCTTCTGGACACGGACCGAGTCCAG 3588

Qy     976 ttcgctccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaacca 1035
           ||||  |||  ||  || || ||||| || || || || |||  |    |||| ||
Db    3587 TTCGGACCAATAGTGGCCCTAAACCCAGCTACGCTGCTCCCTCTA---CCTGAGGAGGGG 3531

Qy    1036 gtgactcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttaca 1095
           ||   ||||| |||| | |  | |||  ||| |     ||     |    || |||||
Db    3530 CTGCAACATGACTGCCTTGACATCTTGGCTGAAGCCCACGGAACTAGACCAGATCTTACG 3471

Qy    1096 gacataccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtg 1155
           |||   || ||  | |  |  |  | |||||||  ||| || || |||||||    ||
Db    3470 GACCAGCCTCTCCCAGACGCTGACCACACCTGGTACACAGATGGGAGCAGCTTCCTGCAA 3411

Qy    1156 gaaggtaagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagc 1215
           || ||  || | | ||| || || |||||       || ||||       |||||||||
Db    3410 GAGGGGCAGCGCAAGGCCGGAGCAGCAGTAACCACCGAGACCGAGGTAGTCTGGGCCAAA 3351

Qy    1216 agcctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttg 1275
           |||||  |  || || || || ||||  |||||  | || || |||| || ||||| ||
Db    3350 GCACTGCCAGCCGGGACATCGGCCCAAAGAGCTGAGTTGATAGCGCTCACCCAAGCCTTA 3291

Qy    1276 cggctggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgact 1335
           ||||  ||||| ||     | ||  |||| || || ||| | ||||| ||||| |||
Db    3290 AAAATGGCAGAAGGTAAGAAGCTGAATGTTTACACCGATAGCCGTTATGCTTTTGCCACT 3231

Qy    1336 gcacacgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaa 1395
           || ||  | ||||| |  || |||| |    | |||||||| || |||| ||| |  |||
Db    3230 GCCCATATTCACGGAGAAATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAA 3171

Qy    1396 ataaagaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggcta 1455
           || || || || || || ||  |    |||| |  | ||  |      |||| ||||| ||
Db    3170 ATCAAAAATAAGGACGAGATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGACTT 3111

Qy    1456 gctattatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccag 1515
           || || || || || |||||||||||| |  ||           | || || |||  |
Db    3110 AGCATAATTCATTGCCCGGGACATCAGAAGGGAAACCGCGCGGAGGCAAGGGGCAACAGG 3051

Qy    1516 atggctgaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaa 1575
           ||||| ||||  | |||    | | ||||     |  |  |    | ||   | |
Db    3050 ATGGCCGACCAAGCGGCCCGAGAAGTAGCCACTAGAGAAACTCCAGAGACTTCCACACTT 2991

Qy    1576 acgcccaaagccccagaaccc-agacgacagtacaccctagaagactggcaagagataaa 1634
           |    ||   ||| ||| | ||    | |||   |        ||  |  || | |
```

```
Db   2990 CTGATAGAAAATTCAGCCCCCTATACTCATGAACATTTTCACTATACGGTGACTGACATA 2931

Qy   1635 aaagatagaccagttctctgagactc--------cggaggggacctgctatacctcatat 1686
          ||||||   |   ||  | | | |         |    ||  || |     |   |
Db   2930 AAAGATCTGACTAAACTAGGGGCCACTTATGACGATGCAAAGAAGTGTTGGGTTTATCAG 2871

Qy   1687 gggaaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatcgtcta 1746
          || |||    |  |||| |  ||      || |||     |  |     | ||||   |
Db   2870 GGAAAGCCTGTAATGCCTGATCAATTCACCTTTGAACTATTAGATTTTCTTCATCAATTG 2811

Qy   1747 acccacctagg------aactaaacacctgcagcagttggtcagaacatcccctttatcat 1800
          ||||||||  |      ||  | | |    |  |         |     |  |||||| |
Db   2810 ACCCACCTCAGTTTCTCAAAAACAAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTAC 2751

Qy   1801 gttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctg 1860
          | ||||  |     ||  |    |    |    |   |||    ||||    | |
Db   2750 ATGCTGAACCGGGATCGAACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAG 2691

Qy   1861 gttaatgctaatccttccagaatacctccaggaaagagactaaggggaagccacccaggc 1920
          || ||||| |     ||     |    | ||| |  ||| |  | ||   ||  || |||
Db   2690 GTCAATGCCAGCAAGTCTGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCGACCCGGC 2631

Qy   1921 gctcactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatcta 1980
          | |||||||||| | || ||||||||||||||| || |     || ||  | ||||||||
Db   2630 ACCCACTGGGAAATTGATTTCACTGAGGTAAAACCTGGCCTGTATGGGTATAAATATCTT 2571

Qy   1981 ttggtttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagact 2040
          || |||||  |||||| || || |||||||||| ||||  || || |||||||| |||
Db   2570 TTAGTTTTCATAGACACTTTCTCTGGATGGGTAGAAGCTTTCCCAACCAAGAAAGAAACT 2511

Qy   2041 tcaaccgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaag 2100
          | |  || ||  | |||||  |||| || ||||| || || ||||| || || ||  ||
Db   2510 GCCAAAGTTGTAACCAAGAAGCTACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAG 2451

Qy   2101 gtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaag 2160
          ||| | ||  | |||||||| || || |||||  || ||||||||||  | | ||| |
Db   2450 GTATTGGGAACCGACAATGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGAT 2391

Qy   2161 atattggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggta 2220
          |||||||| |||||||||||||| |||||||| |||||||||| || ||||| ||||||
Db   2390 TTATTGGGGGTTGATTGGAAACTACATTGTGCTTACAGACCCCAGAGTTCAGGTCAGGTA 2331

Qy   2221 gagaggatgaatagaaccattaaagagacccttaccaaattgaccacagagactggcatt 2280
          || || |||||||| || || || |||||  | || ||||||||     |  ||||||  |
Db   2330 GAAAGAATGAATAGGACAATCAAGGAGACTTTAACTAAATTGACGCTTGCAACTGGCTCT 2271

Qy   2281 aatgattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacag 2340
          |  || ||| ||   || || ||| |||  || | | | |   | ||||| || || |
Db   2270 AGGGACTGGGTGCTCCTGCTTCCCCTTGCCCTGTATCGAGCCCGCAACACGCCGGGCCCC 2211

Qy   2341 tttgggctgacccccctatgaattgctctacgggggaccccccccgttggcagaaattgcc 2400
          .  || || ||||| |||||| |  | || |||| |||||| ||  | | | |  |  |
Db   2210 CACGGTCTCACCCCATATGAAATCTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCT 2151

Qy   2401 tttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctc 2460
            | | | |  ||  | ||    | |  ||   || |     ||    |  |||| |||
Db   2150 GATCCTGACATGGCAAAGGTTACTCATAACCCCTCTCTCCAAGCTCATTTACAGGCACTC 2091

Qy   2461 gagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcagg------agga 2514
          |  |||   |||   ||  ||||  | | |   ||  || ||| ||       ||
Db   2090 TACCTGGTCCAGCACGAAGTCTGGAGACCGTTGGCGGCAGCTTACCAAGAGCAACTAGAC 2031

Qy   2515 gacttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgt 2574
          |  ||| || || |  |||| || || ||  |||| |  ||  |  | ||||
Db   2030 CGGCCGGTAGTACCTCACCCTTTCCGGGTCGGTGACACAGTGTGGGTCCGCAGACACCAA 1971

Qy   2575 gcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacg 2634
          |   ||| || ||  | || ||||| |||| ||| ||| ||  |||| || || ||
Db   1970 ACTAAAAATCTAGAACCCCGCTGGAAAGGACCCTATACCGTCCTACTGACTACCCCCACC 1911

Qy   2635 gctgtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccggc---- 2690
          ||| | ||||| || || ||  |  | ||||||| ||  ||||||| ||| | ||
Db   1910 GCTCTCAAAGTAGACGGCATTGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGAC 1851

Qy   2691 ----------gccacctcccgattcggggtggaaagccgaaaagactgaaaatccccctt 2739
                     ||||||  | || |||   |||   |   ||    |  ||||||||||
Db   1850 ACCAGGATTGAGCCACCAGCAGAATCGACATGGCGTGTTCAACGCTCCCAAAATCCCCTA 1791

Qy   2740 aagcttcgcctccatcgcgtggttccttactctgtcaataactcctcaagttaatggtaa 2799
          |||  |  |  |    |||| |    | || ||      |   ||| ||  ||   |  |
Db   1790 AAGATAAGATTGACCCGCGGGACCTCCTAATCCCCTTAATTCTCTTCCTGTCTCTCAAAG 1731

Qy   2800 acgccttgtggacagccc 2817
             |||   |   |||| |
Db   1730 GGGCCAGATCCGCAGCAC 1713
```

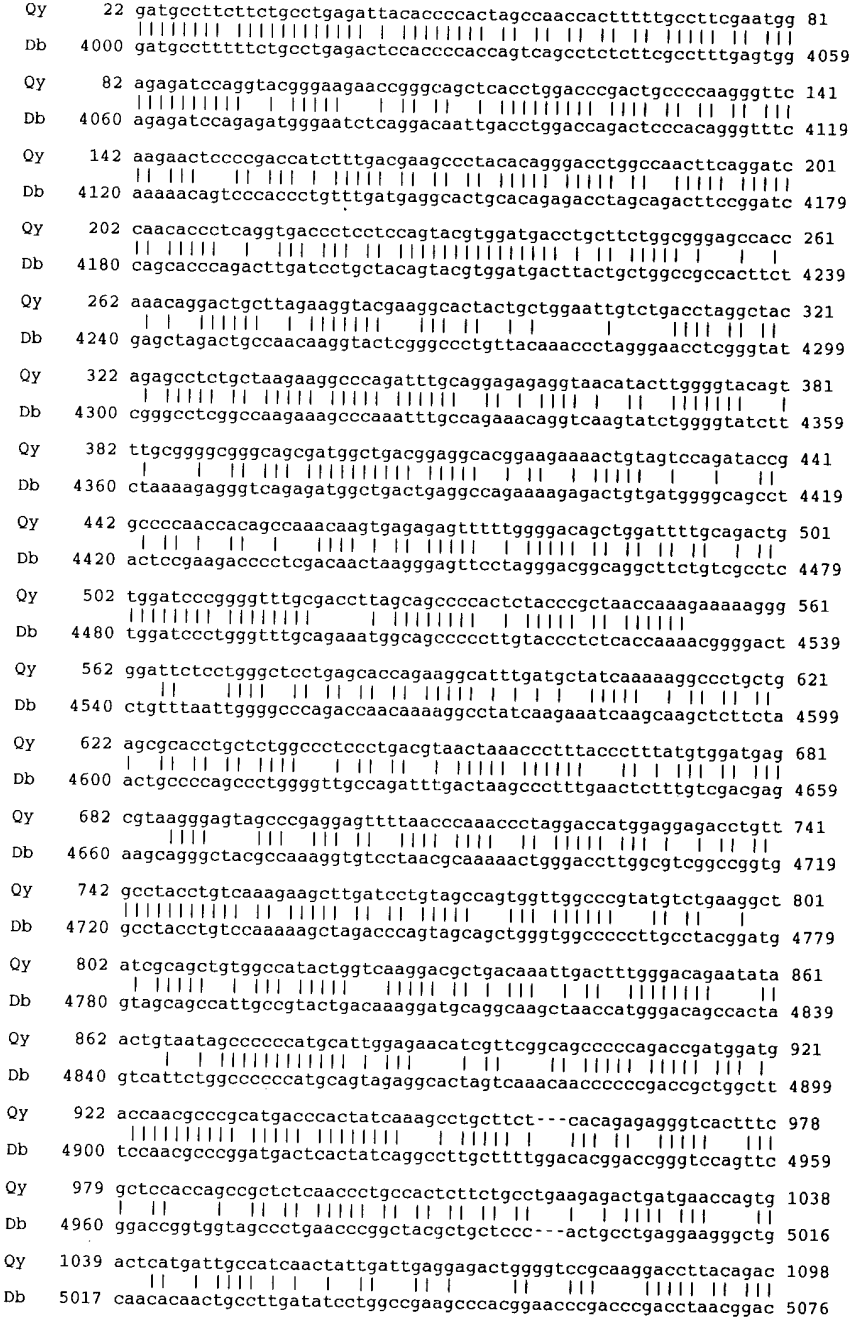
```
RESULT  12
PCT-US93-08041-9/c
; Sequence 9, Application PC/TUS9308041
;  GENERAL INFORMATION:
;    APPLICANT:  The Public Health Research Institute of the City of
;    APPLICANT:  New York, Inc.
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08041
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/938,100
;      FILING DATE:  28-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/_____
;      FILING DATE:  20-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-08041-9


  Query Match             28.1%;  Score 932;  DB 5;  Length 10367;
  Best Local Similarity   59.9%;  Pred. No. 4.3e-273;
  Matches 1701;  Conservative    0;  Mismatches 1095;  Indels   42;  Gaps    7;

Qy      19 acagatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaa 78
           | |||||| || |||||||||||| | |||||||| || ||  | || || ||||| ||
Db    4547 AAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCAGTCAGTCTCTCTTCGCCTTTGAG 4488

Qy      79 tggagagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaaggg 138
           ||||||||||||  | |||||   | || ||  | ||||||||| |||| || || ||
```


Db 4487 TGGAGAGATCCAGAGATGGGAATCTCAGGACAATTAACCTGGACCAGACTCCCGCAGGGT 4428
Qy 139 ttcaagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcagg 198
Db 4427 TTCAAAAACAGTCCCACCCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGG 4368
Qy 199 atccaacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagcc 258
Db 4367 ATCCAGCACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCCACT 4308
Qy 259 accaaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggc 318
Db 4307 TCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCTAGGGGACCTCGGA 4248
Qy 319 tacagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtac 378
Db 4247 TATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTAT 4188
Qy 379 agtttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagata 438
Db 4187 CTTCTAAAAGAGGGTCAGAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAG 4128
Qy 439 ccggccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcaga 498
Db 4127 CCTACTCCGAAGACCCCTCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGC 4068
Qy 499 ctgtggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaa 558
Db 4067 CTCTGGATCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACGGGG 4008
Qy 559 gggggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctg 618
Db 4007 ACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGATCAAGCAGGCTCTC 3948
Qy 619 ctgagcgcacctgctctggccctccctgacgtaactaaaccctttacccttatgtggat 678
Db 3947 TTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTAAGCCCTTCGAACTTTTTGTTGAC 3888
Qy 679 gagcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacct 738
Db 3887 GAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAAACTGGGGCCTTGGCGTCGGCCG 3828
Qy 739 gttgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaag 798
Db 3827 GTGGCCTACCTGTCCAAAAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGG 3768
Qy 799 gctatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaat 858
Db 3767 ATGGTAGCAGCCATCGCCGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAGCCA 3708
Qy 859 ataactgtaatagcccccatgcattggagaacatcgttcggcagcccccagaccgatgg 918
Db 3707 CTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTTAAGCAACCCCCTGATCGCTGG 3648
Qy 919 atgaccaacgcccgcatgacccactatcaaagcctgcttct---cacagagagggtcact 975
Db 3647 CTCTCCAACGCCCGAATGACCCACTACCAGGCTCTGCTTCTGGACACGGACCGAGTCCAG 3588
Qy 976 ttcgctccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaacca 1035
Db 3587 TTCGGACCAATAGTGGCCCTAAACCCAGCTACGCTGCTCCCTCTA---CCTGAGGAGGGG 3531
Qy 1036 gtgactcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttaca 1095
Db 3530 CTGCAACATGACTGCCTTGACATCTTGGCTGAAGCCCACGGAACTAGACCAGATCTTACG 3471
Qy 1096 gacataccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtg 1155
Db 3470 GACCAGCCTCTCCCAGACGCTGACCACACCTGGTACACAGATGGGAGCAGCTTCCTGCAA 3411
Qy 1156 gaaggtaagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagc 1215
Db 3410 GAGGGGCAGCGCAAGGCCGGAGCAGCAGTAACCACCGAGACCGAGGTAGTCTGGGCCAAA 3351
Qy 1216 agcctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttg 1275
Db 3350 GCACTGCCAGCCGGGACATCGGCCCAAAGAGCTGAGTTGATAGCGCTCACCCAAGCCTTA 3291
Qy 1276 cggctggccgaagggaaatccataaacatttatacggacagcaggtatgccctttgcgact 1335
Db 3290 AAAATGGCAGAAGGTAAGAAGCTGAATGTTTACACCGATAGCCGTTATGCTTTTGCCACT 3231
Qy 1336 gcacacgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaa 1395
Db 3230 GCCCATATTCACGGAGAAATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAA 3171
Qy 1396 ataaagaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggcta 1455
Db 3170 ATCAAAAATAAGGACGAGATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGACTT 3111
Qy 1456 gctattatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccag 1515
Db 3110 AGCATAATTCATTGCCCGGGACATCAGAAGGGAAACCGCGCGGAGGCAAGGGGCAACAGG 3051
Qy 1516 atggctgaccgggttgccaagcaggcagcccaggctgttaaccttctgcctataatagaa 1575
Db 3050 ATGGCCGACCAAGCGGCCCGAGAAGTAGCCACTAGAGAAACTCCAGAGACTTCCACACTT 2991
Qy 1576 acgcccaaagccccagaaccc-agacgacagtacaccctagaagactggcaagagataaa 1634
Db 2990 CTGATAGAAAATTCAGCCCCCTATACTCATGAACATTTTCACTATACGGTGACTGACATA 2931
Qy 1635 aaagatagaccagttctctgagactc--------cggaggggacctgctatacctcatat 1686
Db 2930 AAAGATCTGACTAAACTAGGGGCCACTTATGACGATGCAAAGAAGTGTTGGGTTTATCAG 2871
Qy 1687 gggaaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatcgtcta 1746
Db 2870 GGAAAGCCTGTAATGCCTGATCAATTCACCTTTGAACTATTAGATTTTCTTCATCAATTG 2811
Qy 1747 acccacctagg------aactaaacacctgcagcagttggtcagaacatcccctttatcat 1800
Db 2810 ACCCACCTCAGTTTCTCAAAAACAAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTAC 2751
Qy 1801 gttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctg 1860
Db 2750 ATGCTGAACCGGGATCGAACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAG 2691
Qy 1861 gttaatgctaatccttccagaatacctccaggaaagagactaaggggaagccacccaggc 1920
Db 2690 GTCAATGCCAGCAAGTCTGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCGACCCGGC 2631
Qy 1921 gctcactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatcta 1980
Db 2630 ACCCACTGGGAAATTGATTTCACTGAGGTAAAACCTGGCCTGTATGGGTATAAATATCTT 2571
Qy 1981 ttggtttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagact 2040
Db 2570 TTAGTTTTCATAGACACTTTCTCTGGATGGGTAGAAGCTTTCCCAACCAAGAAAGAAACT 2511
Qy 2041 tcaaccgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaag 2100
Db 2510 GCCAAAGTTGTAACCAAGAAGCTACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAG 2451
Qy 2101 gtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaag 2160
Db 2450 GTATTGGGAACCGACAATGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGAT 2391
Qy 2161 atattggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggta 2220
Db 2390 TTATTGGGGGTTGATTGGAAACTACATTGTGCTTACAGACCCCAGAGTTCAGGTCAGGTA 2331
Qy 2221 gagaggatgaatagaaccattaaagagaccccttaccaaattgaccacagagactggcatt 2280
Db 2330 GAAAGAATGAATAGGACAATCAAGGAGACTTTAACTAAATTGACGCTTGCAACTGGCTCT 2271



RESULT 13
US-09-011-745-3
; Sequence 3, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 7308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
US-09-011-745-3

Query Match 27.6%; Score 916.2; DB 4; Length 7308;
Best Local Similarity 60.6%; Pred. No. 2.2e-268;
Matches 1633; Conservative 0; Mismatches 1033; Indels 27; Gaps 7;

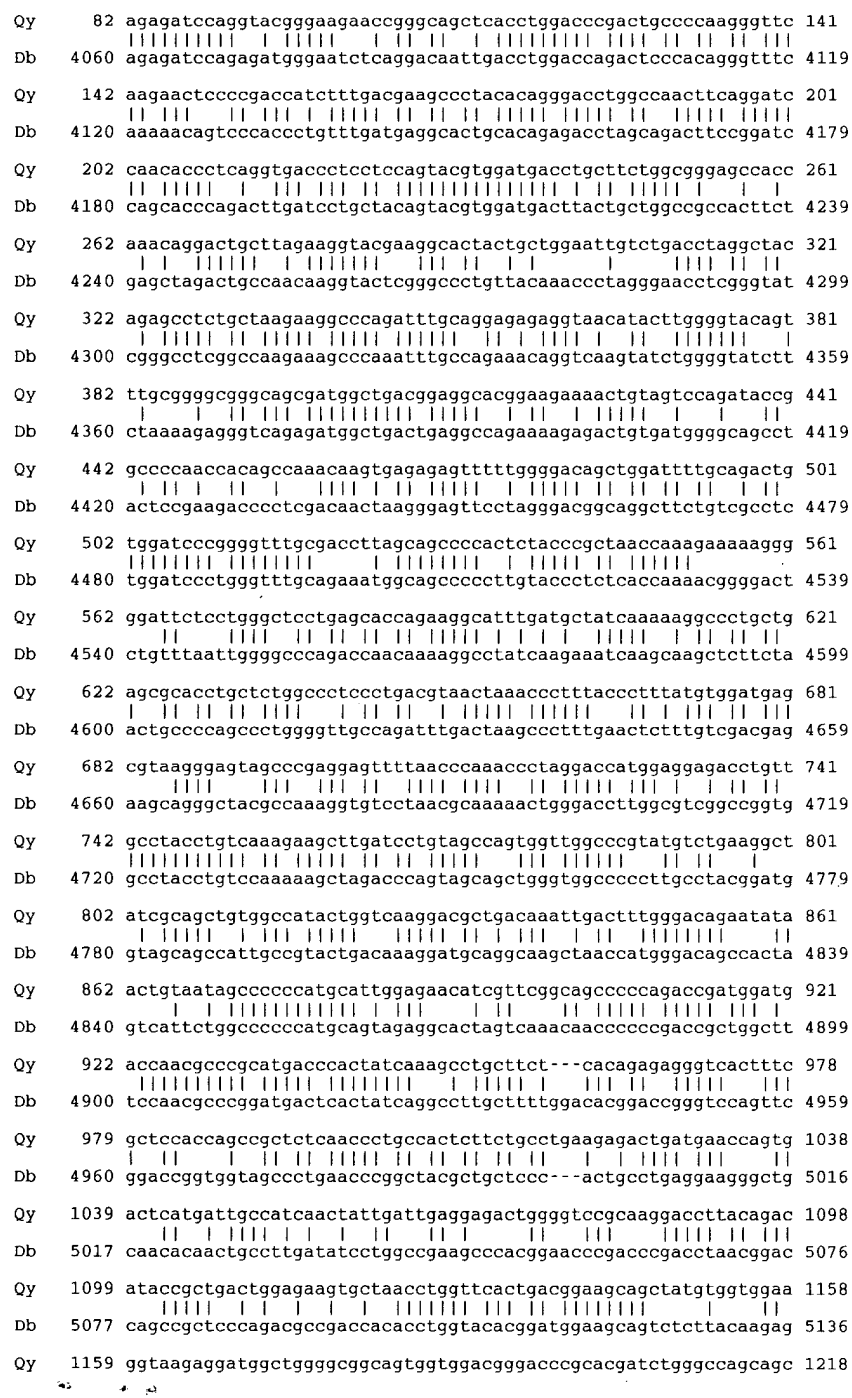


```
Qy   1099 ataccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaa 1158
Db   5077 cagccgctcccagacgccgaccacacctggtacacggatggaagcagtctcttacaagag 5136

Qy   1159 ggtaagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagc 1218
Db   5137 ggacagcgtaaggcgggagctgcggtgaccaccgagaccgaggtaatctgggctaaagcc 5196

Qy   1219 ctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcgg 1278
Db   5197 ctgccagccgggacatccgctcagcgggctgaactgatagcactcacccaggccctaaag 5256

Qy   1279 ctggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgca 1338
Db   5257 atggcagaaggtaagaagctaaatgtttatactgatagccgttatgcttttgctactgcc 5316

Qy   1339 cacgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaata 1398
Db   5317 catatccatggagaaatatacagaaggcgtgggttgctcacatcagaaggcaaagagatc 5376

Qy   1399 aagaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagct 1458
Db   5377 aaaaataaagacgagatcttggccctactaaaagccctctttctgcccaaaagacttagc 5436

Qy   1459 attatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatg 1518
Db   5437 ataatccattgtccaggacatcaaaagggacacagcgccgaggctagaggcaaccggatg 5496

Qy   1519 gctgaccgggttgccaagcaggcagcccaggctgtta-accttctgcctataatagaaac 1577
Db   5497 gctgaccaagcggcccgaaaggcagccatcacagagactccagacacctctaccctcctc 5556

Qy   1578 gcccaaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaa 1637
Db   5557 atagaaaattcatcaccctacacctcagaacattttcattacacagtgactgatataaag 5616

Qy   1638 gatagacc--------agttctctgagactccggaggggacctgctatacctcatatggg 1689
Db   5617 gacctaaccaagttgggggccatttatgataaaacaaagaagtattgggtctaccaagga 5676

Qy   1690 aaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatcgtctaacc 1749
Db   5677 aaacctgtgatgcctgaccagtttacttttgaattattagactttcttcatcagctgact 5736

Qy   1750 cacct-aggaactaaacacctgcagcagttggtcagaacatccccttatcatgttctgag 1808
Db   5737 cacctcagcttctcaaaaatgaaggctctcctagagagaagccacagtccctactacatg 5796

Qy   1809 gctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgc 1868
Db   5797 ctgaaccgggatcgaacactcaaaaatatcactgagacctgcaaagcttgtgcacaagtc 5856

Qy   1869 taatccttccagaatacctccaggaaagagactaagg-----ggaagccacccaggcgct 1923
Db   5857 aacgccagcaagtctgccgttaaacagggaactagggtccgcgggcatcggcccggcact 5916

Qy   1924 cactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatctattg 1983
Db   5917 cattgggagatcgatttcaccgagataaagcccggattgtatggctataaatatcttcta 5976

Qy   1984 gtttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagacttca 2043
Db   5977 gtttttatagatacctttttctggctggatagaagccttcccaaccaagaaagaaaccgcc 6036

Qy   2044 accgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaaggta 2103
Db   6037 aaggtcgtaaccaagaagctactagaggagatcttccccaggttcggcatgcctcaggta 6096

Qy   2104 atagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaagata 2163
Db   6097 ttgggaactgacaatgggcctgccttcgtctccaaggtgagtcagacagtggccgatctg 6156

Qy   2164 ttggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggtagag 2223
Db   6157 ttggggattgattggaaattacattgtgcatacagaccccaaagctcaggccaggtagaa 6216

Qy   2224 aggatgaatagaaccattaaagagacccttaccaaattgaccacagagactggcattaat 2283
Db   6217 agaatgaatagaaccatcaaggagactttaactaaattaacgcttgcaactggctctaga 6276

Qy   2284 gattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacagttt 2343
Db   6277 gactgggtgctcctactcccttagccctgtaccgagcccgcaacacgccgggcccccat 6336

Qy   2344 gggctgacccccctatgaattgctctacgggggacccccccgttggcagaaattgccttt 2403
Db   6337 ggcctcaccccatatgagatcttatatgggqcacccccgcccctttgtaaacttccctgac 6396

Qy   2404 gcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcgag 2463
Db   6397 cctgacatgacaagagttactaacagcccctctctccaagctcacttacaggctctctac 6456

Qy   2464 tgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagacttgcaa 2523
Db   6457 ttagtccagcacgaagtctggagacctctggcggcagcctaccaagaacaactggaccga 6516

Qy   2524 ------gttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgca 2577
Db   6517 ccggtggtacctcaccctaccgagtcggcgacacagtgtgggtccgccgacaccagact 6576

Qy   2578 ggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacggct 2637
Db   6577 aagaacctagaacctcgctggaaaggaccttacacagtcctgctgaccacccccaccgcc 6636

Qy   2638 gtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccggc 2690
Db   6637 ctcaaagtagacggcatcgcagcttggatacacgccgcccacgtgaaggctgc 6689
```

RESULT 14
US-09-011-745-4
; Sequence 4, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 7308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
US-09-011-745-4

Query Match 27.6%; Score 916.2; DB 4; Length 7308;
Best Local Similarity 60.6%; Pred. No. 2.2e-268;
Matches 1633; Conservative 0; Mismatches 1033; Indels 27; Gaps 7;

```
Qy     22 gatgccttcttctgcctgagattacaccccactagccaaccacttttttgccttcgaatgg 81
Db   4000 gatgccttttttctgcctgagactccaccccaccagtcagcctctcttcgcctttgagtgg 4059
```


Qy 82 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 141
Db 4060 agagatccagagatgggaatctcaggacaattgacctggaccagactcccacagggtttc 4119
Qy 142 aagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatc 201
Db 4120 aaaaacagtcccaccctgtttgatgaggcactgcacagagacctagcagacttccggatc 4179
Qy 202 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 261
Db 4180 cagcacccagacttgatcctgctacagtacgtggatgacttactgctggccgccacttct 4239
Qy 262 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 321
Db 4240 gagctagactgccaacaaggtactcgggccctgttacaaaccctagggaacctcgggtat 4299
Qy 322 agagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagt 381
Db 4300 cgggcctcggccaagaaagcccaaatttgccagaaacaggtcaagtatctggggtatctt 4359
Qy 382 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 441
Db 4360 ctaaaagagggtcagagatggctgactgaggccagaaaagagactgtgatggggcagcct 4419
Qy 442 gccccaaccacagccaaacaagtgagagagtttttggggacagctggattttgcagactg 501
Db 4420 actccgaagacccctcgacaactaagggagttcctagggacggcaggcttctgtcgcctc 4479
Qy 502 tggatcccgggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 561
Db 4480 tggatccctgggtttgcagaaatggcagccccccttgtaccctctcaccaaaacggggact 4539
Qy 562 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 621
Db 4540 ctgtttaattggggcccagaccaacaaaaggcctatcaagaaatcaagcaagctcttcta 4599
Qy 622 agcgcacctgctctggccctccctgacgtaactaaacccttttaccctttatgtggatgag 681
Db 4600 actgccccagccctggggttgccagatttgactaagccctttgaactctttgtcgacgag 4659
Qy 682 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 741
Db 4660 aagcagggctacgccaaaggtgtcctaacgcaaaaactgggaccttggcgtcggccggtg 4719
Qy 742 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 801
Db 4720 gcctacctgtccaaaaagctagacccagtagcagctgggtggccccctgcctacggatg 4779
Qy 802 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatata 861
Db 4780 gtagcagccattgccgtactgacaaaggatgcaggcaagctaaccatgggacagccacta 4839
Qy 862 actgtaatagccccccatgcattggagaacatcgttcggcagcccccagaccgatggatg 921
Db 4840 gtcattctggcccccatgcagtagaggcactagtcaaacaacccccgaccgctggctt 4899
Qy 922 accaacgcccgcatgacccactatcaaagcctgcttct---cacagagagggtcactttc 978
Db 4900 tccaacgcccggatgactcactatcaggccttgcttttggacacggaccgggtccagttc 4959
Qy 979 gctccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtg 1038
Db 4960 ggaccggtggtagccctgaacccggctacgctgctccc---actgcctgaggaagggctg 5016
Qy 1039 actcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagac 1098
Db 5017 caacacaactgccttgatatcctggccgaagcccacggaacccgacccgacctaacggac 5076
Qy 1099 ataccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaa 1158
Db 5077 cagccgctcccagacgccgaccacacctggtacacggatggaagcagtctcttacaagag 5136
Qy 1159 ggtaagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagc 1218
Db 5137 ggacagcgtaaggcgggagctgcggtgaccaccgagaccgaggtaatctgggctaaagcc 5196
Qy 1219 ctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcgg 1278
Db 5197 ctgccagccgggacatccgctcagcgggctgaactgatagcactcacccaggccctaaag 5256
Qy 1279 ctggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgca 1338
Db 5257 atggcagaaggtaagaagctaaatgtttatactgatagccgttatgcttttgctactgcc 5316
Qy 1339 cacgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaata 1398
Db 5317 catatccatggagaaatatacagaaggcgtgggttgctcacatcagaaggcaaagagatc 5376
Qy 1399 aagaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagct 1458
Db 5377 aaaaataaagacgagatcttggccctactaaaagccctctttctgcccaaaagacttagc 5436
Qy 1459 attatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatg 1518
Db 5437 ataatccattgtccaggacatcaaaagggacacagcgccgaggctagaggcaaccggatg 5496
Qy 1519 gctgaccgggttgccaagcaggcagcccaggctgtta-accttctgcctataatagaaac 1577
Db 5497 gctgaccaagcggcccgaaaggcagccatcacagagactccagacacctctaccctcctc 5556
Qy 1578 gcccaaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaa 1637
Db 5557 atagaaaattcatcaccctacacctcagaacattttcattacacagtgactgatataaag 5616
Qy 1638 gatagacc--------agttctctgagactccggaggggacctgctatacctcatatggg 1689
Db 5617 gacctaaccaagttgggggccatttatgataaaacaaagaagtattgggtctaccaagga 5676
Qy 1690 aaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatcgtctaacc 1749
Db 5677 aaacctgtgatgcctgaccagtttacttttgaattattagactttcttcatcagctgact 5736
Qy 1750 cacct-aggaactaaacacctgcagcagttggtcagaacatccccttatcatgttctgag 1808
Db 5737 cacctcagcttctcaaaaatgaaggctctcctagagagaagccacagtccctactacatg 5796
Qy 1809 gctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgc 1868
Db 5797 ctgaaccgggatcgaacactcaaaaatatcactgagacctgcaaagcttgtgcacaagtc 5856
Qy 1869 taatccttccagaatacctccaggaaagagactaagg-----ggaagccacccaggcgct 1923
Db 5857 aacgccagcaagtctgccgttaaacagggaactagggtccgcgggcatcggcccggcact 5916
Qy 1924 cactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatctattg 1983
Db 5917 cattgggagatcgatttcaccgagataaagcccggattgtatggctataaatatcttcta 5976
Qy 1984 gtttttgtagacacctttcaggatgggtagaggcttatcctactaagaaagagacttca 2043
Db 5977 gtttttatagatacctttctggctggatagaagccttcccaaccaagaaagaaaccgcc 6036
Qy 2044 accgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaaggta 2103
Db 6037 aaggtcgtaaccaagaagctactagaggagatcttccccaggttcggcatgcctcaggta 6096
Qy 2104 atagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaagata 2163
Db 6097 ttgggaactgacaatgggcctgccttcgtctccaaggtgagtcagacagtggccgatctg 6156
Qy 2164 ttggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggtagag 2223
Db 6157 ttggggattgattggaaattacattgtgcatacagaccccaaagctcaggccaggtagaa 6216
Qy 2224 aggatgaatagaaccattaaagagacccttaccaaattgaccacagagactggcattaat 2283

```
Db   6217 agaatgaatagaaccatcaaggagactttaactaaattaacgcttgcaactggctctaga 6276

Qy   2284 gattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacagttt 2343
          || ||| ||   || || ||||| |  || |   | |   | ||||| || || |    |
Db   6277 gactgggtgctcctactcccctagccctgtaccgagcccgcaacacgccgggcccccat 6336

Qy   2344 gggctgacccccctatgaattgctctacgggggaccccccccgttggcagaaattgccttt 2403
          || || ||||| |||||  |  | || |||| |||||| ||  | | | |  |  |
Db   6337 ggcctcaccccatatgagatcttatatggggcacccccgccccttgtaaacttccctgac 6396

Qy   2404 gcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcgag 2463
            |  | |   |    ||         ||   || |     ||    |  |||| ||| |
Db   6397 cctgacatgacaagagttactaacagcccctctctccaagctcacttacaggctctctac 6456

Qy   2464 tgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagacttgcaa 2523
          |  ||   |||   ||  ||||  |  ||   ||  ||||||  || |  |      | |
Db   6457 ttagtccagcacgaagtctggagacctctggcggcagcctaccaagaacaactggaccga 6516

Qy   2524 ------gttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgca 2577
                ||  ||  || |  | || ||| || ||  |||| |  ||  | || ||||   |
Db   6517 ccggtggtacctcacccttaccgagtcggcgacacagtgtgggtccgccgacaccagact 6576

Qy   2578 ggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacggct 2637
            ||||| ||  |||| ||||| ||||||||    || ||  ||||||| || || ||
Db   6577 aagaacctagaacctcgctggaaaggaccttacacagtcctgctgaccacccccaccgcc 6636

Qy   2638 gtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccggc 2690
            | ||||| || || ||| |  | ||||| || ||  ||||||| ||| | ||
Db   6637 ctcaaagtagacggcatcgcagcttggatacacgccgcccacgtgaaggctgc 6689
```

RESULT 15
US-09-011-745-2
; Sequence 2, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 7616
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
US-09-011-745-2

Query Match 27.6%; Score 916.2; DB 4; Length 7616;
Best Local Similarity 60.6%; Pred. No. 2.3e-268;
Matches 1633; Conservative 0; Mismatches 1033; Indels 27; Gaps 7;

```
Qy     22 gatgccttcttctgcctgagattacaccccactagccaaccacttttgccttcgaatgg 81
          |||||||| |||||||||||| | |||||||| || || || || || ||||| || |||
Db   3161 gatgccttttttctgcctgagactccaccccaccagtcagcctctcttcgcctttgagtgg 3220

Qy     82 agagatccaggtacgggaagaaccgggcagctcacctggacccgactgccccaagggttc 141
          ||||||||||  | |||||  | || ||  | ||||||||| |||| || || || |||
Db   3221 agagatccagagatgggaatctcaggacaattgacctggaccagactcccacagggtttc 3280

Qy    142 aagaactccccgaccatctttgacgaagccctacacagggacctggccaacttcaggatc 201
          || |||   || ||| | ||||| || || || ||||| ||||| ||  ||||| |||||
Db   3281 aaaaacagtcccaccctgtttgatgaggcactgcacagagacctagcagacttccggatc 3340

Qy    202 caacaccctcaggtgaccctcctccagtacgtggatgacctgcttctggcgggagccacc 261
          || |||||  |  ||| ||| || |||||||||||||||| | || ||||| |   |  |
Db   3341 cagcacccagacttgatcctgctacagtacgtggatgacttactgctggccgccacttct 3400

Qy    262 aaacaggactgcttagaaggtacgaaggcactactgctggaattgtctgacctaggctac 321
          | |  ||||||  | |||||||   ||| ||  | |       |     |||| || ||
Db   3401 gagctagactgccaacaaggtactcgggccctgttacaaaccctagggaacctcgggtat 3460

Qy    322 agagcctctgctaagaaggcccagatttgcaggagaggaggtaacatacttggggtacagt 381
          | ||||| || ||||| ||||| ||||||  || | |||| |  ||  |||||||   |
Db   3461 cgggcctcggccaagaaagcccaaatttgccagaaacaggtcaagtatctggggtatctt 3520

Qy    382 ttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccg 441
          |    |  || ||| |||||||||| |||||  | ||  | |||||  |   |   ||
Db   3521 ctaaaagagggtcagagatggctgactgaggccagaaaagagactgtgatggggcagcct 3580

Qy    442 gccccaaccacagccaaacaagtgagagagtttttggggacagctggatttttgcagactg 501
          | || |  ||  |   |||| | || |||||  | ||||| || || || ||  | ||
Db   3581 actccgaagacccctcgacaactaagggagttcctagggacggcaggcttctgtcgcctc 3640

Qy    502 tggatcccgggggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggg 561
          |||||||| ||||||||     | ||||||||  | ||||| || ||||||
Db   3641 tggatccctgggtttgcagaaatggcagccccccttgtaccctctcaccaaaacggggact 3700

Qy    562 ggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctg 621
            ||    ||||  || || || || ||||| | | |  |  |||||  | || || ||
Db   3701 ctgtttaattggggcccagaccaacaaaaggcctatcaagaaatcaagcaagctcttcta 3760

Qy    622 agcgcacctgctctggccctccctgacgtaactaaaccctttaccctttatgtggatgag 681
          |  || || || ||||   | || ||  | ||||| ||||||   || | ||| || |||
Db   3761 actgccccagccctggggttgccagatttgactaagccctttgaactctttgtcgacgag 3820

Qy    682 cgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggaggagacctgtt 741
              ||||    |||  ||| ||  |||| ||||  || ||||| ||| |  | || ||
Db   3821 aagcagggctacgccaaaggtgtcctaacgcaaaaactgggaccttggcgtcggccggtg 3880

Qy    742 gcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggct 801
          |||||||||| || ||||| || || |||||   ||| ||||||   || ||   |
Db   3881 gcctacctgtccaaaaagctagacccagtagcagctgggtggccccccttgcctacggatg 3940

Qy    802 atcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatata 861
          | |||||  | ||| |||||   ||||| || | |||  | ||  |||||||||    ||
Db   3941 gtagcagccattgccgtactgacaaaggatgcaggcaagctaaccatgggacagccacta 4000

Qy    862 actgtaatagcccccccatgcattggagaacatcgttcggcagcccccagaccgatggatg 921
            |  | ||||||||||||| | |||    | ||    || ||||| ||||| ||| |
Db   4001 gtcattctggcccccccatgcagtagaggcactagtcaaacaacccccccgaccgctggctt 4060

Qy    922 accaacgcccgcatgacccactatcaaagcctgcttct---cacagagagggtcactttc 978
          |||||||||| ||||| ||||||||    | ||||| |   ||| ||  |||||   |||
Db   4061 tccaacgcccggatgactcactatcaggccttgcttttggacacggaccgggtccagttc 4120

Qy    979 gctccaccagccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtg 1038
          |  ||    |  || || ||||| || || || || ||   |  | |||| |||    ||
Db   4121 ggaccggtggtagccctgaacccggctacgctgctccc---actgcctgaggaagggctg 4177

Qy   1039 actcatgattgccatcaactattgattgaggagactggggtccgcaaggaccttacagac 1098
            ||  | |||| | |  |  ||   || |     ||   |||    ||||| || |||
Db   4178 caacacaactgccttgatatcctggccgaagcccacggaacccgacccgacctaacggac 4237

Qy   1099 ataccgctgactggagaagtgctaacctggttcactgacggaagcagctatgtggtggaa 1158
            |||||  | |  |  |  |  ||||||| ||| || ||||||||     |    ||
Db   4238 cagccgctcccagacgccgaccacacctggtacacggatggaagcagtctcttacaagag 4297

Qy   1159 ggtaagaggatggctggggcggcagtggtggacgggacccgcacgatctgggccagcagc 1218
          ||  || | | ||| || || || |||     || ||||      |||||||| |    |
Db   4298 ggacagcgtaaggcgggagctgcggtgaccaccgagaccgaggtaatctgggctaaagcc 4357

Qy   1219 ctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcgg 1278
          ||||| |  || || || || ||   |||||| || || || ||||| || ||  |   |
```

```
Db   4358 ctgccagccgggacatccgctcagcgggctgaactgatagcactcacccaggccctaaag 4417

Qy   1279 ctggccgaagggaaatccataaacatttatacggacagcaggtatgcctttgcgactgca 1338
          |||| ||||| ||     ||||  ||||||| || ||| | ||||| ||||| |||||
Db   4418 atggcagaaggtaagaagctaaatgtttatactgatagccgttatgcttttgctactgcc 4477

Qy   1339 cacgtacacggggccatctataaacaaaggggttgcttacctcagcagggagggaaata 1398
          ||  | || || |  || || | |    | |||||||| || |||| ||| |  || ||
Db   4478 catatccatggagaaatatacagaaggcgtgggttgctcacatcagaaggcaaagagatc 4537

Qy   1399 aagaacaaagaggaaattctaagcctattagaagccttacatttgccaaaaaggctagct 1458
          || || ||||| || ||  |   |||| || ||||| |   | |||| ||||| ||
Db   4538 aaaaataaagacgagatcttggccctactaaaagccctctttctgcccaaaagacttagc 4597

Qy   1459 attatacactgtcctggacatcagaaagccaaagatctcatatctagagggaaccagatg 1518
          || || || ||||| |||||||| || |   |      |    |||||||| |||| ||||
Db   4598 ataatccattgtccaggacatcaaaagggacacagcgccgaggctagaggcaaccggatg 4657

Qy   1519 gctgaccgggttgccaagcaggcagcccaggctgtta-accttctgcctataatagaaac 1577
          |||||||  |  |||    |||||||||    | |  |  ||     ||| ||       |
Db   4658 gctgaccaagcggcccgaaaggcagccatcacagagactccagacacctctaccctcctc 4717

Qy   1578 gcccaaagccccagaacccagacgacagtacaccctagaagactggcaagagataaaaaa 1637
               |||   |   | || | ||  ||| |||   |          |  |  |||| |||
Db   4718 atagaaaattcatcaccctacacctcagaacattttcattacacagtgactgatataaag 4777

Qy   1638 gatagacc--------agttctctgagactccggaggggacctgctatacctcatatggg 1689
          ||   | |         |  |  | |  |        ||  |  |    ||   | ||
Db   4778 gacctaaccaagttgggggccatttatgataaaacaaagaagtattgggtctaccaagga 4837

Qy   1690 aaggaaatcctgccccacaaagaagggttagaatatgtccaacagatacatcgtctaacc 1749
          ||     |  ||||  || |       || ||||    |  |     | ||||  || ||
Db   4838 aaacctgtgatgcctgaccagtttacttttgaattattagactttcttcatcagctgact 4897

Qy   1750 cacct-aggaactaaacacctgcagcagttggtcagaacatccccttatcatgttctgag 1808
          |||||  ||   ||  || |      ||  |     |||  | ||  |    | |  |    |
Db   4898 cacctcagcttctcaaaaatgaaggctctcctagagagaagccacagtccctactacatg 4957

Qy   1809 gctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgc 1868
             | | ||  | |   ||||     |   ||   |      |    |   ||   || |
Db   4958 ctgaaccgggatcgaacactcaaaaatatcactgagacctgcaaagcttgtgcacaagtc 5017

Qy   1869 taatccttccagaatacctccaggaaagagactaagg-----ggaagccacccaggcgct 1923
           |  ||  | ||  | ||   |  | |  |||| ||     ||    |  || ||| ||
Db   5018 aacgccagcaagtctgccgttaaacagggaactagggtccgcgggcatcggcccggcact 5077

Qy   1924 cactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatctattg 1983
          || |||||  | || ||||| ||| |||||||  |     || ||  | ||||||||  |
Db   5078 cattgggagatcgatttcaccgagataaagcccggattgtatggctataaatatcttcta 5137

Qy   1984 gtttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagacttca 2043
          |||||| |||| |||||||| || ||| |||| || |  || || |||||||| ||  |
Db   5138 gtttttatagatacctttttctggctggatagaagccttcccaaccaagaaagaaaccgcc 5197

Qy   2044 accgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaaggta 2103
          |  || ||  | |||||  |||| ||||| || || || || || || || ||| |||||
Db   5198 aaggtcgtaaccaagaagctactagaggagatcttccccaggttcggcatgcctcaggta 5257

Qy   2104 atagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaagata 2163
           | ||  | |||||||| || || |||||  || |||| ||||||  | ||||| |  |
Db   5258 ttgggaactgacaatgggcctgccttcgtctccaaggtgagtcagacagtggccgatctg 5317

Qy   2164 ttggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggtagag 2223
          |||||||||||||||||| | ||||||||||||||||||||||||||||||| |||||||
Db   5318 ttggggattgattggaaattacattgtgcatacagaccccaaagctcaggccaggtagaa 5377

Qy   2224 aggatgaatagaaccattaaagagaccttaccaaattgaccacagagactggcattaat 2283
          || |||||||||||||| || |||||  | || ||||| ||    |  ||||||  ||
Db   5378 agaatgaatagaaccatcaaggagactttaactaaattaacgcttgcaactggctctaga 5437

Qy   2284 gattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacagttt 2343
          || ||| ||   || || ||||| |  || |   | |   | ||||| || || |    |
Db   5438 gactgggtgctcctactccccttagccctgtaccgagcccgcaacacgccgggccccat 5497

Qy   2344 gggctgaccccctatgaattgctctacgggggaccccccccgttggcagaaattgccttt 2403
          || || ||||| |||||  |  | || |||| |||||| ||  | | | |  |  |
Db   5498 ggcctcaccccatatgagatcttatatggggcaccccgccccttgtaaacttccctgac 5557

Qy   2404 gcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcgag 2463
           | | |   |  ||         ||   || |     ||   |  |||| ||| |
Db   5558 cctgacatgacaagagttactaacagcccctctctccaagctcacttacaggctctctac 5617

Qy   2464 tgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagacttgcaa 2523
          |  ||   |||   ||  ||||  |  ||   ||  ||||||  || |  |      | |
Db   5618 ttagtccagcacgaagtctggagacctctggcggcagcctaccaagaacaactggaccga 5677

Qy   2524 ------gttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgca 2577
                ||  || || |  | || ||| || ||  |||| |  ||  | || ||||   |
Db   5678 ccggtggtacctcacccttaccgagtcggcgacacagtgtgggtccgccgacaccagact 5737

Qy   2578 ggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacggct 2637
             ||||| ||  |||| |||||| ||||||||    || ||  |||||||| || || ||
Db   5738 aagaacctagaacctcgctggaaaggaccttacacagtcctgctgaccacccccaccgcc 5797

Qy   2638 gtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccggc 2690
           | ||||| || || ||| |  | ||||| || ||  ||||||| ||| | ||
Db   5798 ctcaaagtagacggcatcgcagcttggatacacgccgcccacgtgaaggctgc 5850
```

Search completed: February 24, 2002, 02:39:30
Job time: 20121 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 02:39:30 ; Search time 212.72 Seconds
(without alignments)
8739.930 Million cell updates/sec

Title: US-09-171-553B-3
Perfect score: 8209
Sequence: 1 gtggtgtacgactgtgggcc.....aaaaaaaaaaaaaaaaaaa 8209

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	6282.4	76.5	8132	4	US-08-766-528-3	Sequence 3, Appli
	2	5489.8	66.9	7333	4	US-08-766-528-2	Sequence 2, Appli
	3	4986	60.7	8060	4	US-08-766-528-1	Sequence 1, Appli
	4	3376.6	41.1	3482	3	US-09-111-085-3	Sequence 3, Appli
	5	2295.8	28.0	8655	3	US-09-075-272-1	Sequence 1, Appli
	6	2244.4	27.3	8535	3	US-08-716-351A-1	Sequence 1, Appli
	7	1747	21.3	8323	1	US-08-110-300A-8	Sequence 8, Appli
	8	1747	21.3	8323	2	US-08-886-642-8	Sequence 8, Appli
	9	1747	21.3	8323	5	PCT-US93-08041-8	Sequence 8, Appli
c	10	1738.8	21.2	8202	1	US-08-258-420-13	Sequence 13, Appl
	11	1687.4	20.6	8332	3	US-08-850-961-1	Sequence 1, Appli
c	12	1596	19.4	10367	1	US-08-110-300A-9	Sequence 9, Appli
c	13	1596	19.4	10367	2	US-08-886-642-9	Sequence 9, Appli
c	14	1596	19.4	10367	5	PCT-US93-08041-9	Sequence 9, Appli
	15	1568.4	19.1	2462	3	US-09-111-085-1	Sequence 1, Appli
	16	1550.4	18.9	7308	4	US-09-011-745-3	Sequence 3, Appli
	17	1550.4	18.9	7616	4	US-09-011-745-2	Sequence 2, Appli
	18	1545.6	18.8	7308	4	US-09-011-745-4	Sequence 4, Appli
	19	961	11.7	2000	4	US-09-376-781-25	Sequence 25, Appl
	20	961	11.7	2000	4	US-09-376-781-30	Sequence 30, Appl
	21	852.6	10.4	3674	1	US-08-105-483-324	Sequence 324, App
	22	852.6	10.4	3674	1	US-08-709-209-324	Sequence 324, App
	23	852.6	10.4	3674	1	US-08-458-101-324	Sequence 324, App
	24	815.4	9.9	6363	2	US-08-929-967-6	Sequence 6, Appli
	25	705.4	8.6	1493	4	US-09-376-781-24	Sequence 24, Appl
	26	617.4	7.5	6312	4	US-09-011-745-7	Sequence 7, Appli
	27	510.8	6.2	1140	4	US-09-603-185-5	Sequence 5, Appli
	28	359.4	4.4	10970	3	US-08-716-351A-5	Sequence 5, Appli
	29	331.6	4.0	698	4	US-09-376-781-18	Sequence 18, Appl
	30	307	3.7	6028	4	US-09-011-745-5	Sequence 5, Appli
	31	303.8	3.7	1965	1	US-08-258-420-9	Sequence 9, Appli
	32	301.4	3.7	2499	1	US-08-105-483-310	Sequence 310, App
	33	301.4	3.7	2499	1	US-08-709-209-310	Sequence 310, App
	34	301.4	3.7	2499	1	US-08-458-101-310	Sequence 310, App
	35	300.6	3.7	2001	3	US-08-850-961-13	Sequence 13, Appl
	36	298.8	3.6	949	5	PCT-US91-08254-1	Sequence 1, Appli
c	37	298.8	3.6	949	5	PCT-US91-08254-2	Sequence 2, Appli
	38	298.2	3.6	1911	1	US-08-258-420-8	Sequence 8, Appli
	39	296.8	3.6	1979	4	US-08-552-369-19	Sequence 19, Appl
	40	293.4	3.6	3925	4	US-09-011-745-9	Sequence 9, Appli
	41	290.2	3.5	1914	1	US-08-258-420-7	Sequence 7, Appli
	42	290	3.5	6061	4	US-09-011-745-6	Sequence 6, Appli
	43	278	3.4	278	4	US-09-376-781-10	Sequence 10, Appl
	44	274.8	3.3	789	1	US-08-181-335B-3	Sequence 3, Appli
	45	274.8	3.3	789	1	US-08-181-335B-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-3

Query Match 76.5%; Score 6282.4; DB 4; Length 8132;
Best Local Similarity 87.1%; Pred. No. 0;
Matches 7165; Conservative 0; Mismatches 931; Indels 134; Gaps 19;

Qy 1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 3 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 62
Qy 61 aagaccgcttctcgtgagtgatt-tggggtgtcgcctcttccgagcccggacgaggggga 119
Db 63 AAGACCGCTTCTCGTGAGTGATTAAGGGGAGTCGCCTTTTCCGAGCCTGGAGG------- 115
Qy 120 ttgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctt 179
Db 116 --TTCTTTTTGCTGGTCTTACATTTGGGGGCTCGTCCGGGATCTGTCGCGGCCACCCCTA 173
Qy 180 acacccgagaaccgacttggaggta-aagggatccccttttggaacgtgtgtg--tgtgtc 236
Db 174 ACACCCGAGAACCGACTTGGAGGTAAAAAGGATCCTCTTTTTAACGTGTATGCATGTACC 233
Qy 237 ggccggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctg 296
Db 234 GGCCGGCGTCTCTGTTCTGAGTGTCTGTTTTCAGTGGTGCGCGCTTTCGGTTTGCAGCTG 293
Qy 297 tcctctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacag 356
Db 294 TCCTCTCAGGCCGTAAGGGCTGGGGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAG 353
Qy 357 gctgccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcc 416
Db 354 GCTGCTGCCCTGGGGGACGCCCCGGGAGGTGAGGAGAGCCAGGGACGCCTGGTGGTCTCC 413
Qy 417 tactgtcggtcagaggaccgagttctgttgttgaagcgaaagcttccccctccgcggccg 476
Db 414 TACTGTCGGTCAGAGGACCGAATTCTGTTGCTGAAGCGAAAGCTTCCCCCTCCGCGACCG 473
Qy 477 tccgactcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttg 536
Db 474 TCCGACTCTTTTGCCTGCTTGTGGAAGACGTGGACGGGTCACGTGTGTCTGGATCTGTTG 533
Qy 537 gtttctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacag 596
Db 534 GTTTCTGTTTTGTGTGTCTTTGTCTTGTGTGTCCTTGTCTACAGTTTTAATATGGGACAG 593
Qy 597 acagtgactacccccctagtttgactctcgaccattggactgaagttagatccagggct 656
Db 594 ACGGTGACGACCCCTCTTAGTTTGACTCTCGACCATTGGACTGAAGTTAAATCCAGGGCT 653
Qy 657 cataatttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatgg 716
Db 654 CATAATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGTCTCTGAATGG 713
Qy 717 ccaacattcgatgttggatggccatcagaggggacctttaattctgaaattatcctggct 776
Db 714 CCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCTGAGATTATCCTGGCT 773
Qy 777 gttaaggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatc 836
Db 774 GTTAAAGCAGTTATTTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAGCCCTATATC 833
Qy 837 cttacgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaa 896
Db 834 CTTACGTGGCAAGATTTGGCAGAGGATCCTCCGCCATGGGTTAAACCATGGCTGAATAAG 893
Qy 897 ccaagaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaa 956
Db 894 CCAAGAAAGCCAGGTCCCCGAATTCTGGCTCTTGGAGAGAAAAACAAACACTCGGCTGAA 953
Qy 957 aaagtcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccgga 1016
Db 954 AAAGTCAAGCCCTC-TCCTCATATCTACCCCGAGATTGAGGAGCCACCGGCTTGGCCGGA 1012
Qy 1017 accccaacctgttcccccaccccctatccagcacagggtgctgtgagggga-cctctgc 1075
Db 1013 ACCCCAATCTGTTCCCCCACCCCCTTATCTGGCACAGGGTGCCGCGAGGGGACCCTTTGC 1072
Qy 1076 ccctcctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgc 1135
Db 1073 CCCTCCTGGAGCTCCGGCGGTGGAGGGACCTGCTGCAGGGACTCGGAGCCGGAGGGGCGC 1132
Qy 1136 caccccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccat 1195
Db 1133 CACCCCGGAGCGGACAGACGAGATCGCGACATTACCGCTGCGCACGTACGGCCCTCCCAC 1192
Qy 1196 gccaggggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataa 1255
Db 1193 ACCGGGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAA 1252
Qy 1256 ttggaaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtgga 1315
Db 1253 TTGGAAAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGA 1312
Qy 1316 gtcccttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacact 1375
Db 1313 GTCCCTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACT 1372
Qy 1376 cttcacaaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggc 1435
Db 1373 CTTCACAACCGAGGAGCGAGAGAGAATTCTATTAGAGGCTAGAAAAAATGTTCCTGGGGC 1432
Qy 1436 cgacgggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccc 1495
Db 1433 CGACGGGCGACCCACGCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTAACTCGCCC 1492
Qy 1496 cggttgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctct 1555
Db 1493 CGGTTGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCT 1552
Qy 1556 ggtggcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggt 1615
Db 1553 GGTGGCGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAAGT 1612
Qy 1616 gatgcagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcag 1675
Db 1613 GATGCAGGGACCGAATGAACCCCCCTCTGTTTTTCTTGAGAGGCTCTTGGAAGCCTTCAG 1672
Qy 1676 gcggttcacccctttttgatcctacctcagaggcccagaaagcctcagtggccctggcctt 1735
Db 1673 GCGGTACACCCCTTTTGATCCCACCTCAGAGGCCCAAAAAGCCTCAGTGGCTTTGGCCTT 1732
Qy 1736 cattgggcagtcggctctggatatcaggaagaaacttcagagactggaaagggttacagga 1795
Db 1733 TATAGGACAGTCAGCCTTGGATATTAGAAAGAAGCTTCAGAGACTGGAAGGGTTACAGGA 1792
Qy 1796 ggctgagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacaga 1855
Db 1793 GGCTGAGTTACGTGATCTAGTGAAGGAGGCAGAGAAAGTATATTACAAAAGGGAGACAGA 1852
Qy 1856 agaggagaaggaacagagaaaagaaaaggagagagaagaaagggaggaaaagacgtgatag 1915
Db 1853 AGAAGAAAGGGAACAAAGAAAAGAGAGAGAAAGAGAGGAAAGGGAGGAAAGACGTAATAA 1912
Qy 1916 acggcaagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcag 1975
Db 1913 ACGGCAAGAGAAGAATTTGACTAAGATCTTGGCTGCAGTGGTTGAAGGGAAAAGCAATAC 1972
Qy 1976 ggagagagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctggg 2035
Db 1973 GGAAAGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGG 2032
Qy 2036 caataggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggc 2095
Db 2033 CAATAGGACCCCACTCGACAAGGACCAATGTGCATATTGTAAAGAAAGAGGACACTGGGC 2092
Qy 2096 aaggaactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataa 2155
Db 2093 AAGGAACTGCCCCAAGAAGGGAAACAAAGGACCAAGGATCCTAGCTCTAGAAGAAGATAA 2152
Qy 2156 agattaggggagacggggttcggaccccctccccgagcccagggtaactttgaaggtgga 2215
Db 2153 AGATTAGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGA 2212

Qy 2216 ggggcaaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaacc 2275
Db 2213 GGGGCAACCAGTTGAGTTCCTGGTTGATACCGGAGCGAAACATTCAGTGCTACTACAGCC 2272
Qy 2276 attaggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagta 2335
Db 2273 ATTAGGAAAACTAAAAGATAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACAACAGTA 2332
Qy 2336 tccatggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttct 2395
Db 2333 TCCATGGACTACCCGAAGAACAGTTGACTTGGGAGTGGGACGGGTAACCCACTCGTTTCT 2392
Qy 2396 ggtcatccctgagtgcccagtacccttctaggtagagacttactgaccaagatgggagc 2455
Db 2393 GGTCATACCTGAGTGCCCAGCACCCCTCTTAGGTAGAGACTTATTGACCAAGATGGGAGC 2452
Qy 2456 tcaaatttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgt 2515
Db 2453 ACAAATTTCTTTTGAACAAGGGAAACCAGAAGTGTCTGCAAATAACAAACCTATCACTGT 2512
Qy 2516 gttgaccctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatca 2575
Db 2513 GTTGACCCTCCAATTAGATGACGAATATCGACTATACTCTCCCCTAGTAAAGCCTGATCA 2572
Qy 2576 agatatacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatggg 2635
Db 2573 AAATATACAATTCTGGTTGGAACAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGG 2632
Qy 2636 tttggcaaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatc 2695
Db 2633 TTTGGCAAAGCAAGTTCCCCCACAAGTTATTCAACTGAAGGCCAGTGCCACACCAGTGTC 2692
Qy 2696 agtcagacagtaccccttgagtagagaggctcgagaaggaatttggccgcatgttcaaag 2755
Db 2693 AGTCAGACAGTACCCCTTGAGTAAAGAAGCTCAAGAAGGAATTCGGCCGCATGTCCAAAG 2752
Qy 2756 attaatccaacagggcatcctagttcctgtccaatccccttggaatactcccctgctacc 2815
Db 2753 ATTAATCCAACAGGGCATCCTAGTTCCTGTCCAATCTCCCTGGAATACTCCCCTGCTACC 2812
Qy 2816 ggttaggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataa 2875
Db 2813 GGTTAGAAAGCCTGGGACTAATGACTATCGACCAGTACAGGACTTGAGAGAGGTCAATAA 2872
Qy 2876 aagggtgcaggacatacacccaacggtcccgaacccttataacctcttgagcgccctccc 2935
Db 2873 ACGGGTGCAGGATATACACCCAACAGTCCCGAACCCTTATAACCTCTTGTGTGCTCTCCC 2932
Qy 2936 gcctgaacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagatt 2995
Db 2933 ACCCCAACGGAGCTGGTATACAGTATTGGACTTAAAGGATGCCTTCTTCTGCCTGAGATT 2992
Qy 2996 acaccccactagccaaccactttttgccttcgaatggagagatccaggtacgggaagaac 3055
Db 2993 ACACCCCACTAGCCAACCACTTTTTGCCTTCGAATGGAGAGATCCAGGTACGGGAAGAAC 3052
Qy 3056 cgggcagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttga 3115
Db 3053 CGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGA 3112
Qy 3116 cgaagccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcct 3175
Db 3113 CGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATCCAACACCCTCAGGTGACCCTCCT 3172
Qy 3176 ccagtacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtac 3235
Db 3173 CCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGCAC 3232
Qy 3236 gaaggcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggccca 3295
Db 3233 GAAGGCACTACTGCTGGAATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCA 3292
Qy 3296 gatttgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggct 3355
Db 3293 GATTTGCAGGAGAGAGGTAACATACTTGGGGTACAGTTTGCGGGACGGGCAGCGATGGCT 3352
Qy 3356 gacggaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagt 3415
Db 3353 GACGGAGGCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAAT 3412
Qy 3416 gagagagtttttggggacagctggattttgcagactgtggatcccggggtttgcgacctt 3475
Db 3413 GAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTT 3472
Qy 3476 agcagccccactctacccgctaaccaaagaaaaaggggattctcctgggctcctgagca 3535
Db 3473 AGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGGGAATTCTCCTGGGCTCCTGAGCA 3532
Qy 3536 ccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctccc 3595
Db 3533 CCAGAAGGCATTTGATGCTATCAAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCC 3592
Qy 3596 tgacgtaactaaacccttttaccctttatgtggatgagcgtaagggagtagcccgaggagt 3655
Db 3593 TGACGTAACTAAACCCTTTACCCTTTATGTGGATGAGCGTAAGGGAGTAGCCCGGGGAGT 3652
Qy 3656 tttaacccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagcttga 3715
Db 3653 TTTAACCCAAACCCTAGGACCATGGAGAAGACCTGTCGCCTACCTGTCAAAGAAGCTCGA 3712
Qy 3716 tcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccatactggt 3775
Db 3713 TCCTGTAGCCAGTGGTTGGCCCATATGCCTGAAGGCTATCGCAGCTGTGGCCATACTGGT 3772
Qy 3776 caaggacgctgacaaattgactttgggacagaatataactgtaatagcccccccatgcatt 3835
Db 3773 CAAGGACGCTGACAAATTGACTTTGGGACAGAATATAACTGTAATAGCCCCCCATGCATT 3832
Qy 3836 ggagaacatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacccacta 3895
Db 3833 GGAGAACATCGTTCGGCAGCCCCCAGACCGATGGATGACCAACGCCCGCATGACCCACTA 3892
Qy 3896 tcaaagcctgcttctcacagagagggtcactttcgctccaccagccgctctcaaccctgc 3955
Db 3893 TCAAAGCCTGCTTCTCACAGAGAGGGTCACGTTCGCTCCACCAGCCGCTCTCAACCCTGC 3952
Qy 3956 cactcttctgcctgaagagactgatgaaccagtgactcatgattgccatcaactattgat 4015
Db 3953 CACTCTTCTGCCTGAAGAGACTGATGAACCAGTGACTCATGATTGCCATCAACTATTGAT 4012
Qy 4016 tgaggagactggggtccgcaaggaccttacagacataccgctgactggagaagtgctaac 4075
Db 4013 TGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGGAGAAGTGCTAAC 4072
Qy 4076 ctggttcactgacggaagcagctatgtggtggaaggtaagaggatggctggggcggcagt 4135
Db 4073 CTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGGTAAGAGGATGGCTGGGGCGGCGGT 4132
Qy 4136 ggtggacgggacccgcacgatctgggccagcagcctgccggaaggaacttcagcgcaaaa 4195
Db 4133 GGTGGACGGGACCCGCACGATCTGGGCCAGCAGCCTGCCGGAAGGAACTTCAGCACAAAA 4192
Qy 4196 ggctgagctcatggccctcacgcaagctttgcggctggccgaagggaaatccataaacat 4255
Db 4193 GGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAAACAT 4252
Qy 4256 ttatacggacagcaggtatgcctttgcgactgcacacgtacacggggccatctataaaca 4315
Db 4253 TTATACGGACAGCAGGTATGCCTTTGCGACTGCACACGTACATGGGGCCATCTATAAACA 4312
Qy 4316 aaggggttgcttacctcagcagggagggaaataaagaacaaagaggaaattctaagcct 4375
Db 4313 AAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAGGAAATTCTAAGCCT 4372
Qy 4376 attagaagccttacatttgccaaaaaggctagctattatacactgtcctggacatcagaa 4435

```
Db  4373 ATTAGAAGCCGTACATTTACCAAAAAGGCTAGCTATTATACACTGTCCTGGACATCAGAA 4432

Qy  4436 agccaaagatctcatatctagagggaaccagatggctgaccgggttgccaagcaggcagc 4495
         ||| |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||
Db  4433 AGCTAAAGATCTCATATCCAGAGGAAACCAGATGGCTGACCGGGTTGCCAAGCAGGCAGC 4492

Qy  4496 ccaggctgttaaccttctgcctataatagaaacgcccaaagccccagaacccagacgaca 4555
         |||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  4493 CCAGGGTGTTAACCTTCTGCCTATAATAGAAATGCCCAAAGCCCCAGAACCCAGACGACA 4552

Qy  4556 gtacaccctagaagactggcaagagataaaaaagatagaccagttctctgagactccgga 4615
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4553 GTACACCCTAGAAGACTGGCAAGAGATAAAAAAGATAGACCAGTTCTCTGAGACTCCGGA 4612

Qy  4616 ggggacctgctatacctcatatgggaaggaaatcctgccccacaaagaagggttagaata 4675
         |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  4613 AGGGACCTGCTATACCTCAGATGGGAAGGAAATCCTGCCCCACAAAGAAGGGTTAGAATA 4672

Qy  4676 tgtccaacagatacatcgtctaacccacctaggaactaaacacctgcagcagttggtcag 4735
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4673 TGTCCAACAGATACATCGTCTAACCCACCTAGGAACTAAACACCTGCAGCAGTTGGTCAG 4732

Qy  4736 aacatccccttatcatgttctgaggctaccaggagtggctgactcggtggtcaaacattg 4795
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4733 AACATCCCCTTATCATGTTCTGAGGCTACCAGGAGTGGCTGACTCGGTGGTCAAACATTG 4792

Qy  4796 tgtgccctgccagctggttaatgctaatccttccagaatacctccaggaaagagactaag 4855
         |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db  4793 TGTGCCCTGCCAGCTGGTTAATGCTAATCCTTCCAGAATGCCTCCAGGGAAGAGACTAAG 4852

Qy  4856 gggaagccacccaggcgctcactgggaagtggacttcactgaggtaaagccggctaaata 4915
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4853 GGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAATA 4912

Qy  4916 cggaaacaaatatctattggtttttgtagacacctttttcaggatgggtagaggcttatcc 4975
         |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  4913 CGGAAACAAATACCTATTGGTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCC 4972

Qy  4976 tactaagaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttccaag 5035
         |||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db  4973 TACTAAGAAAGAGACTTCAACCGTGGTGGCTAAAAAAATACTGGAAGAAATTTTTCCAAG 5032

Qy  5036 atttggaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccaggtaag 5095
         |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db  5033 ATTTGGAATACCTAAGGTAATAGGGTCAGACAATGGTCCAGCTTTTGTTGCCCAGGTAAG 5092

Qy  5096 tcagggactggccaagatattggggattgattggaaactgcattgtgcatacagacccca 5155
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5093 TCAGGGACTGGCCAAGATATTGGGGATTGATTGGAAACTGCATTGTGCATACAGACCCCA 5152

Qy  5156 aagctcaggacaggtagagaggatgaatagaaccattaaagagacccttaccaaattgac 5215
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5153 AAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACCCTTACTAAATTGAC 5212

Qy  5216 cacagagactggcattaatgattggatggctctcctgccctttgtgctttttagggtgag 5275
         | | ||||||||| |||||||||||||| |||||||||||||||||||||||||||||||
Db  5213 CGCGGAGACTGGCGTTAATGATTGGATAGCTCTCCTGCCCTTTGTGCTTTTTAGGGTTAG 5272

Qy  5276 gaacacccctggacagtttgggctgacccctatgaattgctctacgggggaccccccccc 5335
         |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  5273 GAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTACTCTACGGGGGACCCCCCCC 5332

Qy  5336 gttggcagaaattgcctttgcacatagtgctgatgtgctgctttcccagcctttgttctc 5395
         |||| |||||||| | || |||||||||||||| ||||||||||||||||||||||||||
Db  5333 ATTGGTAGAAATTGCTTCTGTACATAGTGCTGACGTGCTGCTTTCCCAGCCTTTGTTCTC 5392

Qy  5396 taggctcaaggcgctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggccta 5455
         |||||||||||| || |||||||||| || |||||||||| ||| ||||||||||||||||
Db  5393 TAGGCTCAAGGCACTTGAGTGGGTGAGACAACGAGCGTGGAGGCAACTCCGGGAGGCCTA 5452

Qy  5456 ctc---aggaggagacttgcaagttccacatcgcttccaagttggagattcagtctatgt 5512
         |||   |||||||||||||||  | |||||||| |||||||| |||||||||||||| ||
Db  5453 CTCAGGAGGAGGAGACTTGCAGATCCCACATCGTTTCCAAGTGGGAGATTCAGTCTACGT 5512
```

```
Qy  5513 tagacgccaccgtgcaggaaacctcgagactcggtggaagggaccttatctcgtacttttt 5572
         |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  5513 TAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGCCCTTATCTCGTACTTTT 5572

Qy  5573 gaccacaccaacggctgtgaaagtcgaaggaatccccacctggatccatgcatcccacgt 5632
         |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  5573 GACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGATCCATGCATCCCACGT 5632

Qy  5633 taagccggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaa 5692
         ||| |||||||||||||||| |||||||||| |||||||||||||| ||||||||||||
Db  5633 TAAACCGGCGCCACCTCCCGATTCGGGGTGGAAAGCCGAAAAGACTGAAAATCCCCTTAA 5692

Qy  5693 gcttcgcctccatcgcctggttccttactctaacaataactccccaggccagtagtaaac 5752
         |||||||||||||||| |||||||||||||  |||||||  | ||| | | | |||  |
Db  5693 GCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTCAATAACCTCTCAGACTAATGGTATGC 5752

Qy  5753 gccttatagacagctcgaaccccccatagacctttatcccttacctggctgattattgacc 5812
         || |   |||||||  |||| |||||| ||| ||||| || |||||| | |||| ||||
Db  5753 GCATAGGAGACAGCCTGAACTCCCATAAACCCTTATCTCTCACCTGGTTAATTACTGACT 5812

Qy  5813 ctgatacgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggc 5872
         | |  || ||| | | | | || | ||||| ||| |  ||||||  ||| ||||||||||
Db  5813 CCGGCACAGGTATTAATATCAACAACACTCAAGGGGAGGCTCCTTTAGGAACCTGGTGGC 5872

Qy  5873 ctgaactgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacc 5932
         |||| ||  |  | ||||||  |||   |||  ||   | | |    | |||| ||  |
Db  5873 CTGATCTATACGTTTGCCTCAGATCAGTTATTCCTAGTCTGA---CCTCACCCCCAGATA 5929

Qy  5933 tagtccgtagttatgggttctattgctgcccaggcacag---agaaagagaaatactgtg 5989
         |  ||| |  | | || || |||   ||||||||  ||    | || |  ||| | || |
Db  5930 TCCTCCATGCTCACGGATTTTATGTTTGCCCAGGACCACCAAATAATGGAAAACATTGCG 5989

Qy  5990 ggggttctggggaatccttctgtaggagatggagctgcgtcacctccaacgatggagact 6049
         |   | |  | || | ||| ||||     ||||| ||| || ||||| || |||||| | |
Db  5990 GAAATCCCAGAGATTTCTTTTGTAAACAATGGAACTGTGTAACCTCTAATGATGGATATT 6049

Qy  6050 ggaaatggccgatctctctccaggaccgggtaaaattctcctttgtcaa---------- 6098
         |||||||||  | |||||  |||||  ||||||  || || | ||||||
Db  6050 GGAAATGGCCAACCTCTCAGCAGGATAGGGTAAGTTTTTCTTATGTCAACACCTATACCA 6109

Qy  6099 -ttccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctccccat 6157
          || || |     || |||   |  ||    |  | |      |  ||||| || |
Db  6110 GCTCTGGACAATTTAATTACCTGACCTGGATTAGAACTGGAAGCCCCAAGTGCTCTCCTT 6169

Qy  6158 cagacttagattatctaaagataagtttcact--gaaaggaaaacaggaaaatattcaaa 6215
         ||||| ||||||| ||||| |||||||||||  ||||||||||| |||||||| | ||
Db  6170 CAGACCTAGATTACCTAAAAATAAGTTTCACTGAGAAAGGAAAACAAGAAAATATCCTAA 6229

Qy  6216 agtggataaatggtatgagctggggaatagttttttattatatggcggg----ggagcag 6271
         | ||| |||||||||||   |||||||| || | ||||      ||||      | |||
Db  6230 AATGGGTAAATGGTATGTCTTGGGGAATGGTATATTATGGAGGCTCGGGTAAACAACCAG 6289

Qy  6272 ggtccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatgg 6331
         | |||| | |||| |||||||| |  ||| |   |  |   || || ||  |||| || |
Db  6290 GCTCCATTCTAACTATTCGCCTCAAAATAAACCAGCTG---GAGCCTCCAATGGCTATAG 6346

Qy  6332 gacccgataaagtactggctgaacaggggccccggccctggagccaccgcataacttgc 6391
         ||||  |||  ||  || | |  ||  | |||||  |||
Db  6347 GACCAAATACGGTCTTGACGGGTCAAAGACCCCCAACCC---------------------- 6385

Qy  6392 cggtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtacca 6451
                                                |  | ||  ||  |   ||||  ||
Db  6386 ----------------------------AAGGACCAGGACCATCCTCTAACATAACTT 6415

Qy  6452 ctggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggac 6511
         ||||||     || ||  |  |    || |   |               ||| |  ||
Db  6416 CTGGATCAGACCCCACTGAGTCTAACAGCACGAC---------------TAAAATGGGGG 6460

Qy  6512 agagactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatg 6571
           | ||| || || |||||||||||||||||| |||||  | ||||||||    || || |
Db  6461 CAAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAGG 6520
```

```
Qy   6572 ccacttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggcta 6631
          | || ||||||||||||||| ||  || | || || || ||||| ||||| || ||||||| 
Db   6521 CTACCTCTTCTTGTTGGCTATGCTTAGCTTCGGGCCCACCTTACTATGAAGGAATGGCTA 6580

Qy   6632 aagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaa 6691
          |  | | ||||||||||||||| ||||| |||||| | ||||| |||||||| |||| || 
Db   6581 GAAGAGGGAAATTCAATGTGACAAAAGAACATAGAGACCAATGCACATGGGGATCCCAAA 6640

Qy   6692 ataagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccat 6751
          |||||||||||| ||||| ||||| || || || || |||||||||| | |||||||| 
Db   6641 ATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGAAAGGTTCCCCCAT 6700

Qy   6752 cccaccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtatt 6811
          |||||||||||||||  |    |||| |  | | | ||  |||| || | ||| ||| 
Db   6701 CCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCAAACCTCTGAGAGTCAATATC 6760

Qy   6812 tagtacctggttataacaggtggtgggcatgcaatactgggttaaccccctgtgtttcca 6871
          | |||||||||||| |||||||||||||||| |||||||| |||||||| |||||||||| 
Db   6761 TGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTGTGTTTCCA 6820

Qy   6872 cctcagtcttcaaccaatccaaagatttctgtgtcatggtccaaatcgtccccgagtgt 6931
          |||  || || |||||| | |||||||| ||  | |||||||||| || ||||||||| 
Db   6821 CCTTGGTTTTTAACCAAACTAAAGATTTTTGCATTATGGTCCAAATTGTTCCCCGAGTGT 6880

Qy   6932 actaccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaa 6991
          | ||| |||| ||  |||  |||||||||||||||  |  || | |||| ||| | 
Db   6881 ATTACTATCCCGAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAAGA 6940

Qy   6992 gagaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtag 7051
          |||||||| |||| || || || ||||| |||||||| | ||   ||| |  || |||| 
Db   6941 GAGAACCCATATCTCTGACACTTGCTGTGATGCTCGGACTTGGAGTGGCAGCAGGTGTAG 7000

Qy   7052 gaacagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagc 7111
          ||||||| |||||||||||| |||| |||||||||||||||||| | ||||||| || | | 
Db   7001 GAACAGGAACAGCTGCCCTGGTCACGGGACCACAGCAGCTAGAAACAGGACTTAGTAACC 7060

Qy   7112 tacatgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaag 7171
          |||||       | |||||||||||| |||| || |  | ||||| || ||||| || | 
Db   7061 TACATCGAATTGTAACAGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCTGGAGG 7120

Qy   7172 agtccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgt 7231
          | ||||| || || || |||||||| || ||||||||  | || || |||||| |  | | 
Db   7121 AATCCCTAACCTCCTTATCTGAAGTAGTCCTACAGAATAGAAGAGGGTTAGATTTATTAT 7180

Qy   7232 ttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatc 7291
          |||||| |||||| || |||||| |||||| || || ||||| || || ||||| |||| 
Db   7181 TTCTAAAAGAAGGAGGATTATGTGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTGGATC 7240

Qy   7292 actcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaa 7351
          | ||||| |||||||||||||||||||||||||||||| ||| ||| |||| |||||||| 
Db   7241 ATTCAGGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAAGGTTGGAGAAGCGTCGAA 7300

Qy   7352 gggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatga 7411
          ||||||  ||  ||   || |||||||||||| |||||||||||||||||||| |||| || 
Db   7301 GGGAAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGGTTCAACAGGTCTCTTTGGTTGG 7360

Qy   7412 ccaccctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttg 7471
          | ||||| |||||||| | || || ||| || |||||||  |||||||||| |||||||| 
Db   7361 CTACCCTACTTTCTGCTTTAACAGGACCCTTAATAGTCCTCCTCCTGTTACTCACAGTTG 7420

Qy   7472 ggccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccaga 7531
          |||| ||  | ||||| | |||  |||||||  |||||||||| | |||||||||||| 
Db   7421 GGCCATGTATTATTAACAAGTTAATTGCCTTCATTAGAGAACGAATAAGTGCAGTCCAGA 7480

Qy   7532 tcatggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctagc 7591
          |||||||||||| |||||||||| |||     |||   | ||| ||| || |||||| 
Db   7481 TCATGGTACTTAGACAACAGTACCAAAGCCCGTCTAGC---AGGGAAGCTGGCCGCTAGC 7537

Qy   7592 cttcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaa 7651
          | |||||||||||||||||||||||||||||| ||||||||||||||||||||||||| 
Db   7538 TCTACCAGTTCTAAGATTAGAACTATTAACAAGAGAAGAAGTGGGGAATGAAAGGATGAA 7597

Qy   7652 aatgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgcccccgaa 7711
          ||| |||||||| |     |||| |    || |      ||  |   | |    ||  | 
Db   7598 AATACAACCTAAGCTAATGAGAAGCTTAAAATTGTTCTGAATTCCAGAGTTTGTTCCTTA 7657

Qy   7712 ttacagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagg 7771
          |    |       || |    |    |  | | ||  | |  |    |  |  |  | | 
Db   7658 TAGGTAAAAGATTAGGTTTTTTGCTGTTTTAAAATATGCGGAAGTAAAATAGGCCCTGAG 7717

Qy   7772 gcctgctatcctggcct-aagtaagataacaggaaatgagttgac----taatcgcttat 7826
          |   | ||  ||| | ||     | | |  |  |||| | ||    ||  |  | | 
Db   7718 TACATGTCTCTAGGCATGAAACTTCTTGAAACTATTTGAGATAACAAGAAAAGGGAGTTT 7777

Qy   7827 ctggattctgtaaaactgactggcaccatagaagaattgattacacattgacagccctag 7886
          ||   | ||     |   ||| |   | | |        | | |||| |     | 
Db   7778 CTAACTGCTTGTTTAGCTTCTGTAAAACTGGTTGCGCCATAAAGATGTTGAAATGTTGAT 7837

Qy   7887 tgacctatctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctcc 7946
            || ||||      |||  |||| | |   ||  ||  |      |   |     | 
Db   7838 ACACATATCTTGGTGACAACATGTCTCCCCCACCCCGAAACATGCGCAAATGTGTAACTC 7897

Qy   7947 ggagctatttttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattg 8006
            |  | |||||| | |||||||||| ||||||||||||||||  |||||||  ||| || 
Db   7898 TAAAACAATTTAAATTAATTGGTCCACGAAGCGCGGGCTCTCGAAGTTTTAAATTGACTG 7957

Qy   8007 gtccatggagcgcgggctctcgatattttaaaatgattggtttgtgacgcacaggctttg 8066
          ||   |                |||||||| |||||||||||||| | || | ||||||| 
Db   7958 GTTTGT---------------GATATTTTGAAATGATTGGTTTGTAAAGCGCGGGCTTTG 8002

Qy   8067 ttgtgaaccccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctg 8126
          |||||||||||||||||||||||||||||| ||| ||||||||||||||||||||||||| 
Db   8003 TTGTGAACCCCATAAAAGCTGTCCCGACTCCACACTCGGGGCCGCAGTCCTCTACCCCTG 8062

Qy   8127 cgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcat 8186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db   8063 CGTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCAT 8122

Qy   8187 caaaaaaaaa 8196
          |||||||||| 
Db   8123 CAAAAAAAAA 8132
```

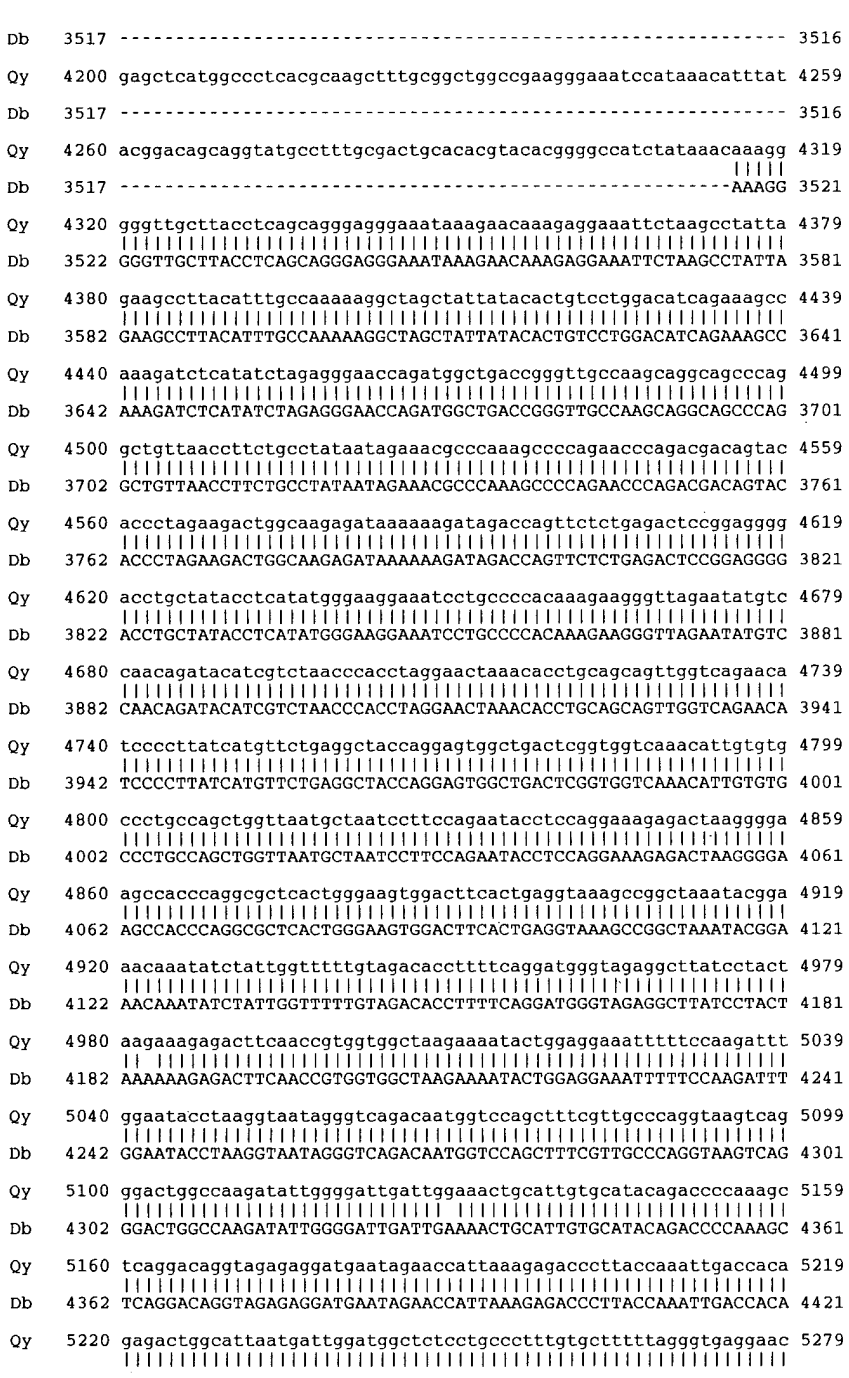
```
RESULT   2
US-08-766-528-2
; Sequence 2, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
```

; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-2

Query Match 66.9%; Score 5489.8; DB 4; Length 7333;
Best Local Similarity 83.0%; Pred. No. 0;
Matches 6856; Conservative 0; Mismatches 402; Indels 1005; Gaps 12;

```
Qy      1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 70

Qy     61 aagaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgaggggggat 120
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     71 AAGACCGCTTCTTGTGAGTGATTTGGGGTGTCGCCTCTTCCGAGCCCGGACGAGGGGGAT 130

Qy    121 tgttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccctta 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 TGTTCTTTTACTGGCCTTTCATTTGGTGCGTTGGCCGGGAAATCCTGCGACCACCCCTTA 190

Qy    181 cacccgagaaccgacttggaggtaaagggatccccttttggaacgtgtgtgtgtgtcggcc 240
          |||||||||||||||||||||||||||||||||||||||||| | ||||||||||||||||
Db    191 CACCCGAGAACCGACTTGGAGGTAAAGGGATCCCCTTTGGAACATATGTGTGTGTCGGCC 250

Qy    241 ggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcct 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 GGCGTCTCTGTTCTGAGTGTCTGTTTTCGGTGATGCGCGCTTTCGGTTTGCAGCTGTCCT 310

Qy    301 ctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 CTCAGACCGTAAGGACTGGAGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAGGCTG 370

Qy    361 ccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctact 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 CCACCCTGGGGGACGCCCCGGGAGGTGGGGAGAGCCAGGGACGCCTGGTGGTCTCCTACT 430

Qy    421 gtcggtcagaggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    431 GTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCG 490

Qy    481 actcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggttt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    491 ACTCTTTTGCCTGCTTGTGGAAGACGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGTTT 550

Qy    541 ctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacag 600
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    551 CTGTTTCGTGTGTCTTTGTCTTGTGCGTCCTTGTCTACAGTTTTAATATGGGACAGACAG 610

Qy    601 tgactacccccctagtttgactctcgaccattggactgaagttagatccagggctcata 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 TGACTACCCCCCTTAGTTTGACTCTCGACCATTGGACTGAAGTTAGATCCAGGGCTCATA 670

Qy    661 atttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 ATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGCCTCTGAATGGCCAA 730

Qy    721 cattcgatgttggatggccatcagaggggacctttaattctgaaattatcctggctgtta 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 CATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCTGAAATTATCCTGGCTGTTA 790

Qy    781 aggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatccttta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 AGGCAATCATTTTTCAGACTGGACCCGGCTCTCATCCTGATCAGGAGCCCTATATCCTTA 850

Qy    841 cgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaa 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    851 CGTGGCAAGATTTGGCAGAAGATCCTCCGCCATGGGTTAAACCATGGCTAAATAAACCAA 910

Qy    901 gaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaaaaag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 GAAAGCCAGGTCCCCGAATCCTGGCTCTTGGAGAGAAAAACAAACACTCGGCCGAAAAAG 970

Qy    961 tcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccggaaccc 1020
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 TCGAGCCCTC-TCCTCGTATCTACCCCGAGATCGAGGAGCCGCCGACTTGGCCGGAACCC 1029

Qy   1021 caacctgttcccccacccccttatccagcacagggtgctgtgagggga-cctctgcccct 1079
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   1030 CAACCTGTTCCCCCACCCCCTTATCCAGCACAGGGTGCTGTGAGGGGACCCTCTGCCCCT 1089

Qy   1080 cctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgccacc 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1090 CCTGGAGCTCCGGTGGTGGAGGGACCTGCTGCCGGGACTCGGAGCCGGAGAGGCGCCACC 1149

Qy   1140 ccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgcca 1199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1150 CCGGAGCGGACAGACGAGATCGCGATATTACCGCTGCGCACCTATGGCCCTCCCATGCCA 1209

Qy   1200 gggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattgg 1259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1210 GGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGG 1269

Qy   1260 aaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtcc 1319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1270 AAAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGAGTCC 1329

Qy   1320 cttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttc 1379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1330 CTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTC 1389

Qy   1380 acaaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgac 1439
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1390 ACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAAAAAAAATGTTCCTGGGGCCGAC 1449

Qy   1440 gggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggt 1499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1450 GGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGATTTCCCTTGACTCGCCCCGGT 1509

Qy   1500 tgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtg 1559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1510 TGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCTGGTG 1569

Qy   1560 gcgggtctccgggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatg 1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1570 GCGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAGGTGATG 1629

Qy   1620 cagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcgg 1679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1630 CAGGGACCGAACGAACCTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGG 1689

Qy   1680 ttcacccctttgatcctacctcagaggcccagaaagcctcagtggccctggccttcatt 1739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1690 TTCACCCCTTTTGATCCTACCTCAGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTCATT 1749

Qy   1740 gggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggct 1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1750 GGGCAGTCGGCTCTGGATATCAGGAAGAAACTTCAGAGACTGGAAGGGTTACAGGAGGCT 1809

Qy   1800 gagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagag 1859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1810 GAGTTACGTGATCTAGTGAGAGAGGCAGAGAAGGTGTATTACAGAAGGGAGACAGAAGAG 1869

Qy   1860 gagaaggaacagagaaaagaaaaggagagagaagaaaggggaggaaagacgtgatagacgg 1919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1870 GAGAAGGAACAGAGAAAAGAAAAGGAGAGAGAAGAAAGGGAGGAAAGACGTGATAGACGG 1929

Qy   1920 caagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggag 1979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1930 CAAGAGAAGAATTTGACTAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAG 1989
```

```
Qy   1980 agagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaat 2039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1990 AGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAAT 2049

Qy   2040 aggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaagg 2099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2050 AGGACCCCACTCGACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGGACACTGGGCAAGG 2109

Qy   2100 aactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataaagat 2159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2110 AACTGCCCCAAGAAGGGAAACAAAGGACCGAAGGTCCTAGCTCTAGAAGAAGATAAAGAT 2169

Qy   2160 taggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtggagggg 2219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2170 TAGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGAGGGG 2229

Qy   2220 caaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaaccatta 2279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2230 CAACCAGTTGAGTTCCTGGTTGATACCGGAGCGGAGCATTCAGTGCTGCTACAACCATTA 2289

Qy   2280 ggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagtatcca 2339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2290 GGAAAACTAAAAGAAAAAAAATCCTGGGTGATGGGTGCCACAGGGCAACGGCAGTATCCA 2349

Qy   2340 tggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttctggtc 2399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2350 TGGACTACCCGAAGAACCGTTGACTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTC 2409

Qy   2400 atccctgagtgcccagtacccctctaggtagagacttactgaccaagatgggagctcaa 2459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2410 ATCCCTGAGTGCCCAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAA 2469

Qy   2460 atttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgtgttg 2519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2470 ATTTCTTTTGAACAAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCACTGTGTTG 2529

Qy   2520 accctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatcaagat 2579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2530 ACCCTCCAATTAGATGATGAATATCGACTATATTCTCCCCAAGTAAAGCCTGATCAAGAT 2589

Qy   2580 atacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatgggtttg 2639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2590 ATACAGTCCTGGTTGGAGCAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGGTTTG 2649

Qy   2640 gcaaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatcagtc 2699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2650 GCAAAGCAAGTTCCCCCACAGGTTATTCAACTGAAGGCCAGTGCTACACCAGTATCAGTC 2709

Qy   2700 agacagtacccccttgagtagagaggctcgagaaggaatttggccgcatgttcaaagatta 2759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2710 AGACAGTACCCCTTGAGTAGAGAGGCTCGAGAAGGAATTTGGCCGCATGTTCAAAGATTA 2769

Qy   2760 atccaacagggcatcctagttcctgtccaatccccttggaatactcccctgctaccggtt 2819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2770 ATCCAACAGGGCATCCTAGTTCCTGTCCAATCCCCTTGGAATACTCCCCTGCTACCGGTT 2829

Qy   2820 aggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataaaagg 2879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2830 AGGAAGCCTGGGACCAATGATTATCGACCAGTACAGGACTTGAGAGAGGTCAATAAAAGG 2889

Qy   2880 gtgcaggacatacacccaacggtcccgaaccccttataacctcttgagcgccctcccgcct 2939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2890 GTGCAGGACATACACCCAACGGTCCCGAACCCTTATAACCTCTTGAGCGCCCTCCCGCCT 2949

Qy   2940 gaacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagattacac 2999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2950 GAACGGAACTGGTACACAGTATTGGACTTAAAAGATGCCTTCTTCTGCCTGAGATTACAC 3009

Qy   3000 cccactagccaaccacttttttgccttcgaatggagagatccaggtacgggaagaaccggg 3059
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   3010 CCCACTAGCCAACCACTTTTTACCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGG 3069

Qy   3060 cagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaa 3119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3070 CAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGACGAA 3129

Qy   3120 gccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccag 3179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3130 GCCCTACACAGGGACCTGGCCAACTTCAGGATCCAACACCCTCAGGTGACCCTCCTCCAG 3189

Qy   3180 tacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaag 3239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3190 TACGTGGATGACCTGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGTACGAAG 3249

Qy   3240 gcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagatt 3299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3250 GCACTACTGCTGGAATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATT 3309

Qy   3300 tgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacg 3359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3310 TGCAGGAGAGAGGTAACATACTTGGGGTACAGTTTGCGGGGCGGGCAGCGATGGCTGACG 3369

Qy   3360 gaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgaga 3419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3370 GAGGCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAGTGAGA 3429

Qy   3420 gagtttttggggacagctggattttgcagactgtggatcccggggtttgcgaccttagca 3479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3430 GAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTTAGCA 3489

Qy   3480 gccccactctacccgctaaccaaagaaaaaggggattctcctgggctcctgagcaccag 3539
          ||||||||||||||||||||||||||
Db   3490 GCCCCACTCTACCCGCTAACCAAAGAA---------------------------------- 3516

Qy   3540 aaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctccctgac 3599
Db   3517 ---------------------------------------------------------- 3516

Qy   3600 gtaactaaaacctttacccttttatgtggatgagcgtaagggagtagcccgaggagtttta 3659
Db   3517 ---------------------------------------------------------- 3516

Qy   3660 acccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagcttgatcct 3719
Db   3517 ---------------------------------------------------------- 3516

Qy   3720 gtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccatactggtcaag 3779
Db   3517 ---------------------------------------------------------- 3516

Qy   3780 gacgctgacaaattgactttgggacagaatataactgtaatagccccccatgcattggag 3839
Db   3517 ---------------------------------------------------------- 3516

Qy   3840 aacatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacccactatcaa 3899
Db   3517 ---------------------------------------------------------- 3516

Qy   3900 agcctgcttctcacagagagggtcactttcgctccaccagccgctctcaaccctgccact 3959
Db   3517 ---------------------------------------------------------- 3516

Qy   3960 cttctgcctgaagagactgatgaaccagtgactcatgattgccatcaactattgattgag 4019
Db   3517 ---------------------------------------------------------- 3516

Qy   4020 gagactggggtccgcaaggaccttacagacataccgctgactggagaagtgctaacctgg 4079
Db   3517 ---------------------------------------------------------- 3516

Qy   4080 ttcactgacggaagcagctatgtggtggaaggtaagaggatggctggggcggcagtggtg 4139
Db   3517 ---------------------------------------------------------- 3516

Qy   4140 gacgggacccgcacgatctgggccagcagcctgccggaaggaacttcagcgcaaaaggct 4199
```


Db 3517 ---------- 3516
Qy 4200 gagctcatggccctcacgcaagctttgcggctggccgaagggaaatccataaacatttat 4259
Db 3517 ---------- 3516
Qy 4260 acggacagcaggtatgcctttgcgactgcacacgtacacggggccatctataaacaaagg 4319
Db 3517 ----------AAAGG 3521
Qy 4320 gggttgcttacctcagcagggagggaaataaagaacaaagaggaaattctaagcctatta 4379
Db 3522 GGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAGGAAATTCTAAGCCTATTA 3581
Qy 4380 gaagccttacatttgccaaaaaggctagctattatacactgtcctggacatcagaaagcc 4439
Db 3582 GAAGCCTTACATTTGCCAAAAAGGCTAGCTATTATACACTGTCCTGGACATCAGAAAGCC 3641
Qy 4440 aaagatctcatatctagagggaaccagatggctgaccgggttgccaagcaggcagcccag 4499
Db 3642 AAAGATCTCATATCTAGAGGGAACCAGATGGCTGACCGGGTTGCCAAGCAGGCAGCCCAG 3701
Qy 4500 gctgttaaccttctgcctataatagaaacgcccaaagccccagaacccagacgacagtac 4559
Db 3702 GCTGTTAACCTTCTGCCTATAATAGAAACGCCCAAAGCCCCAGAACCCAGACGACAGTAC 3761
Qy 4560 accctagaagactggcaagagataaaaaagatagaccagttctctgagactccggagggg 4619
Db 3762 ACCCTAGAAGACTGGCAAGAGATAAAAAAGATAGACCAGTTCTCTGAGACTCCGGAGGGG 3821
Qy 4620 acctgctatacctcatatgggaaggaaatcctgccccacaaagaagggttagaatatgtc 4679
Db 3822 ACCTGCTATACCTCATATGGGAAGGAAATCCTGCCCCACAAAGAAGGGTTAGAATATGTC 3881
Qy 4680 caacagatacatcgtctaacccacctaggaactaaacacctgcagcagttggtcagaaca 4739
Db 3882 CAACAGATACATCGTCTAACCCACCTAGGAACTAAACACCTGCAGCAGTTGGTCAGAACA 3941
Qy 4740 tcccctatcatgttctgaggctaccaggagtggctgactcggtggtcaaacattgtgtg 4799
Db 3942 TCCCCTTATCATGTTCTGAGGCTACCAGGAGTGGCTGACTCGGTGGTCAAACATTGTGTG 4001
Qy 4800 ccctgccagctggttaatgctaatccttccagaatacctccaggaaagagactaagggga 4859
Db 4002 CCCTGCCAGCTGGTTAATGCTAATCCTTCCAGAATACCTCCAGGAAAGAGACTAAGGGGA 4061
Qy 4860 agccacccaggcgctcactgggaagtggacttcactgaggtaaagccggctaaatacgga 4919
Db 4062 AGCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAATACGGA 4121
Qy 4920 aacaaatatctattggtttttgtagacacctttttcaggatgggtagaggcttatcctact 4979
Db 4122 AACAAATATCTATTGGTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACT 4181
Qy 4980 aagaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttccaagattt 5039
Db 4182 AAAAAAGAGACTTCAACCGTGGTGGCTAAGAAAATACTGGAGGAAATTTTTCCAAGATTT 4241
Qy 5040 ggaatacctaaggtaataggtcagacaatggtccagctttcgttgcccaggtaagtcag 5099
Db 4242 GGAATACCTAAGGTAATAGGGTCAGACAATGGTCCAGCTTTCGTTGCCCAGGTAAGTCAG 4301
Qy 5100 ggactggccaagatattggggattgattggaaactgcattgtgcatacagaccccaaagc 5159
Db 4302 GGACTGGCCAAGATATTGGGGATTGATTGAAAACTGCATTGTGCATACAGACCCCAAAGC 4361
Qy 5160 tcaggacaggtagagaggatgaatagaaccattaaagagaccccttaccaaattgaccaca 5219
Db 4362 TCAGGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACCCTTACCAAATTGACCACA 4421
Qy 5220 gagactggcattaatgattggatggctctcctgccctttgtgcttttttagggtgaggaac 5279
Db 4422 GAGACTGGCATTAATGATTGGATGGCTCTCCTGCCCTTTGTGCTTTTTAGGGTGAGGAAC 4481
Qy 5280 acccctggacagtttgggctgacccccctatgaattgctctacgggggaccccccccgttg 5339
Db 4482 ACCCCTGGACAGTTTGGGCTGACCCCCTATAAATTGCTCTACGGGGGACCCCCCCCGTTG 4541
Qy 5340 gcagaaattgcctttgcacatagtgctgatgtgctgctttcccagcctttgttctctagg 5399
Db 4542 GCAGAAATTGCCTTTGCACATAGTGCTGATGTGCTGCTTTCCCAGCCTTTGTTCTCTAGG 4601
Qy 5400 ctcaaggcgctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactca 5459
Db 4602 CTCAAGGCGCTCGAGTGGGTGAGGCAGCGAGCGTGGAAGCAGCTCCGGGAGGCCTACTCA 4661
Qy 5460 ggaggagacttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgc 5519
Db 4662 GGAGGAGACTTGCAAGTTCCACATCGCTTCCAAGTTGGAGATTCAGTCTATGTTAGACGC 4721
Qy 5520 caccgtgcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccaca 5579
Db 4722 CACCGTGCAGGAAACCTCGAGACTCGGTAGAAGGGACCTTATCTCGTACTTTTGACCACA 4781
Qy 5580 ccaacggctgtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccg 5639
Db 4782 CCAACGGCTGTGAAAGTCGAAG---------- 4803
Qy 5640 gcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcgc 5699
Db 4804 ----------GAATCCCCTTAAGCTTCGC 4822
Qy 5700 ctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgccttat 5759
Db 4823 CTCCATCGCGTGGTTCCTTACTCTGTCAATAACTCCTCAAGTTAATGGTAAACGCCTTGT 4882
Qy 5760 agacagctcgaaccccccatagacctttatcccttacctggctgattattgaccctgatac 5819
Db 4883 GGACAGCCCGAACTCCCATAAACCCTTATCTCTCACCTGGTTACTTACTGACTCCGGTAC 4942
Qy 5820 gggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaact 5879
Db 4943 AGGTATTAATATTAACAGCACTCAAGGGGAGGCTCCCTTGGGGACCTGGTGGCCTGAATT 5002
Qy 5880 gcatttctgcctccgattgattaaccccgctgtta------aaagcacacctcccaacct 5933
Db 5003 ATATGTCTGCCTTCGATCAGTAATCCCTGGTCTCAATGACCAGGCCACACCCCCCGATGT 5062
Qy 5934 agtccgtagttatgggttctattgctgcccagg---cacagagaaagagaaatactgtgg 5990
Db 5063 ACTCCGTGCTTACGGGTTTTACGTTTGCCCAGGACCCCCAAATAATGAAGAATATTGTGG 5122
Qy 5991 gggttctggggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactg 6050
Db 5123 AAATCCTCAGGATTTCTTTTGCAAGCAATGGAGCTGCATAACTTCTAATGATGGGAATTG 5182
Qy 6051 gaaatggccgatctctctccaggaccgggtaaaattctcctttgtcaattcc-------- 6102
Db 5183 GAAATGGCCAGTCTCTCAGCAAGACAGAGTAAGTTACTCTTTTGTTAACAATCCTACCAG 5242
Qy 6103 ----------ggcccgggcaagtacaaaa 6121
Db 5243 TTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGATTGGCAACAGCGGGTACAAAA 5302
Qy 6122 tgatgaaactatataaagataagagctgctccccatcagacttagattatctaaagataa 6181
Db 5303 AGATG-TACGAAATAAGCAAATAAGCTGTCATTCGTTAGACCTAGATTACTTAAAAATAA 5361
Qy 6182 gtttcactg--aaaggaaaacaggaaaatattcaaaagtggataaatggtatgagctggg 6239
Db 5362 GTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTGGGTAAATGGTATATCTTGGG 5421
Qy 6240 gaatagtttttat----tatatggcgggggagcagggtccactttaaccattcgcctta 6295
Db 5422 GAATAGTGTACTATGGAGGCTCTGGGAGAAAGAAAGGATCTGTTCTGACTATTCGCCTCA 5481

```
Qy   6296 ggatagagacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaac 6355
          | ||||| ||   ||  ||||| || || || || |||||  |||| |   |||| ||||
Db   5482 GAATAGAAACTCAGATGGAACCTCCGGTTGCTATAGGACCAAATAAGGGTTTGGCCGAAC 5541

Qy   6356 aggggccccccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgc 6415
          | || || ||   ||  ||                                       | |
Db   5542 AAGGACCTCCAATCCAAGA----------------------------------ACAGA 5565

Qy   6416 ggcctgacataacacagccgcctagcaacagtaccactggattgattcctaccaacacgc 6475
          ||||      |  |    |  ||| |  ||| |||||   | || ||  |
Db   5566 GGCCATCTCCTAACCCCTCTGATTACAATACAACCTCTGGATCAGTCCCCACTGA----- 5620

Qy   6476 ctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagg 6535
               ||     |  || |||| |||||   || ||| || || ||||||||||
Db   5621 ----------GCCTAACATCACTATTAAAACAGGGGCGAAACTTTTTAGCCTCATCCAGG 5670

Qy   6536 gagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtc 6595
          ||||||| |||||  | ||||||||    || || || || ||||||||||||||||
Db   5671 GAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTTGCT 5730

Qy   6596 tatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgacca 6655
          || | || || || ||||| |||||||| ||||||| || | | ||||||||||||| |
Db   5731 TAGCTTCGGGCCCACCTTACTATGAGGGAATGGCTAGAGGAGGGAAATTCAATGTGACAA 5790

Qy   6656 aagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagttt 6715
          | || |||||| | |||||||||||||| |||| |||||||||||||| ||||| ||||
Db   5791 AGGAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTACCCTTACTGAGGTTT 5850

Qy   6716 ccgggaaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagta 6775
          | || || || || |||||||| |  | ||||||||||||||||||||||||  |    |
Db   5851 CTGGAAAAGGCACCTGCATAGGGATGGTTCCCCCATCCCACCAACACCTTTGTAACCACA 5910

Qy   6776 ctgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggt 6835
          |||  |  | | | |  |||| || | ||| ||| | ||||||||||| |||||||||
Db   5911 CTGAAGCCTTTAATCGAACCTCTGAGAGTCAATATCTGGTACCTGGTTATGACAGGTGGT 5970

Qy   6836 gggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaag 6895
          ||||||| |||||||| |||||||| ||||||||||||  || |||||||| | ||||
Db   5971 GGGCATGTAATACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAAAG 6030

Qy   6896 atttctgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtcc 6955
          | || || || |||||||||| |||||||| |||||||| |||| ||  |||  ||||
Db   6031 ACTTTTGCGTTATGGTCCAAATTGTCCCCCGGGTGTACTACTATCCCGAAAAAGCAGTCC 6090

Qy   6956 ttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctag 7015
          |||||||||||||||| | ||||| || |||||||||| ||| ||||||| || ||||
Db   6091 TTGATGAATATGACTATAGATATAATCGGCCAAAAAGAGAGCCCATATCCCTGACACTAG 6150

Qy   7016 ctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatca 7075
          |||||||||||||||| ||   ||| |  ||||| |||||||| || |||||||| ||||
Db   6151 CTGTAATGCTCGGATTGGGAGTGGCTGCAGGCGTGGGAACAGGAACGGCTGCCCTAATCA 6210

Qy   7076 caggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatc 7135
          ||||||| || ||||| |||||||||||| || | ||||||       | || |||||||
Db   6211 CAGGACCGCAACAGCTGGAGAAAGGACTTAGTAACCTACATCGAATTGTAACGGAAGATC 6270

Qy   7136 tccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaag 7195
          ||| |||| || |  | ||||| || ||||| || || ||||| || || || |||||||
Db   6271 TCCAAGCCCTAGAAAAATCTGTCAGTAACCTGGAGGAATCCCTAACCTCCTTATCTGAAG 6330

Qy   7196 tggttctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtg 7255
          |||||||||||||| | ||||| ||||||||| | ||||||| |||||| ||||||||||
Db   6331 TGGTTCTACAGAACAGAAGGGGGTTAGATCTGTTATTTCTAAAAGAAGGAGGGTTATGTG 6390

Qy   7256 cagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactcca 7315
          |||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db   6391 TAGCCTTAAAAGAGGAATGCTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCA 6450

Qy   7316 tgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggt 7375
          ||| |||||||||| ||| ||||||||||||||||||||||||||||||||||||||||
Db   6451 TGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTGACCAGGGGT 6510

Qy   7376 ggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacgg 7435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6511 GGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACCACCCTGCTTTCTGCTCTGACGG 6570

Qy   7436 ggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttg 7495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6571 GGCCCCTAGTAGTCCTGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTG 6630

Qy   7496 ttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtacc 7555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6631 TTGCCTTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACC 6690

Qy   7556 aaggccttctgagccaaggagaaactgacctctagccttcccagttctaagattagaact 7615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6691 AAGGCCTTCTGAGCCAAGGAGAAACTGACCTCTAGCCTTCCCAGTTCTAAGATTAGAACT 6750

Qy   7616 attaacaagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaac 7675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6751 ATTAACAAGACAAGAAGTGGGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAAC 6810

Qy   7676 ccaggaagttaataaaaagctctaaatgcccccgaattacagaccctgctggctgccagt 7735
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   6811 CCAGGAAGTTAATAAAAAGCTCTAAATGCCCCCGAATTCCAGACCCTGCTGGCTGCCAGT 6870

Qy   7736 aaataggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaag 7795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6871 AAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTATCCTGGCCTAAGTAAG 6930

Qy   7796 ataacaggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccat 7855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6931 ATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAAAACTGACTGGCACCAT 6990

Qy   7856 agaagaattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactc 7915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6991 AGAAGAATTGATTACACATTGACAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTC 7050

Qy   7916 tgcccaggagcccacgcagatgcggacctccggagctatttaaaatgattggtccacgg 7975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7051 TGCCCAGGAGCCCACGCAGATGCGGACCTCCGGAGCTATTTTAAAATGATTGGTCCACGG 7110

Qy   7976 agcgcgggctctcgatattttaaaatgattggtccatggagcgcgggctctcgatatttt 8035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7111 AGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCATGGAGCGCGGGCTCTCGATATTTT 7170

Qy   8036 aaaatgattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgatt 8095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7171 AAAATGATTGGTTTGTGACGCACAGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATT 7230

Qy   8096 ccgcactcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcg 8155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7231 CCGCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCG 7290

Qy   8156 cttggaataaaaatcctcttgctgtttgcatcaaaaaaaaaaa 8198
          |||||||||||||||||||||||||||||||||||||||||||
Db   7291 CTTGGAATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAAAA 7333
```

```
RESULT   3
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1


  Query Match             60.7%;  Score 4986;  DB 4;  Length 8060;
  Best Local Similarity   95.2%;  Pred. No. 0;
  Matches 5276;  Conservative    0;  Mismatches  245;  Indels   23;  Gaps   12;

Qy      1 gtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2530 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATC 2589

Qy     61 aagaccgcttctcgtgagtgatt-tggggtgtcgcctcttccgagcccggacgaggggga 119
          ||||||||||||||||||||||  |||| |||||||| ||||||||| ||| |
Db   2590 AAGACCGCTTCTCGTGAGTGATTAAGGGGAGTCGCCTTTTCCGAGCCTGGAGG------- 2642

Qy    120 ttgttctttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccaccccctt 179
            |  |||| |||| ||| ||||||| |  | | ||||||  |   ||| ||||||||
Db   2643 --TTCTTTTTGCTGGTCTTACATTTGGGGGCTCGTCCGGGATCTGTCGCGGCCACCCCTA 2700

Qy    180 acacccgagaaccgacttggaggta-aagggatccccctttggaacgtgtgtg--tgtgtc 236
          |||||||||||||||||||||||| || |||||| ||||  |||||||  ||  |||  |
Db   2701 ACACCCGAGAACCGACTTGGAGGTAAAAAGGATCCTCTTTTTAACGTGTATGCATGTACC 2760

Qy    237 ggccggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctg 296
          |||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||||
Db   2761 GGCCGGCGTCTCTGTTCTGAGTGTCTGTTTTCAGTGGTGCGCGCTTTCGGTTTGCAGCTG 2820

Qy    297 tcctctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacag 356
          ||||||||| |||||||| |||| ||||||||||||||||||||||||||||||||||||
Db   2821 TCCTCTCAGGCCGTAAGGGCTGGGGGACTGTGATCAGCAGACGTGCTAGGAGGATCACAG 2880

Qy    357 gctgccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcc 416
          |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 GCTGCTGCCCTGGGGGACGCCCCGGGAGGTGAGGAGAGCCAGGGACGCCTGGTGGTCTCC 2940

Qy    417 tactgtcggtcagaggaccgagttctgttgttgaagcgaaagcttccccctccgcggccg 476
          |||||||||||||||||||| |||||||| ||||||||||||||||||||||||||| |||
Db   2941 TACTGTCGGTCAGAGGACCGAATTCTGTTGCTGAAGCGAAAGCTTCCCCCTCCGCGACCG 3000

Qy    477 tccgactcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttg 536
          |||||||||||||||||||||||||| ||| |||||||| ||||||||||||||||||||
Db   3001 TCCGACTCTTTTGCCTGCTTGTGGAATACGTGGACGGGTCACGTGTGTCTGGATCTGTTG 3060

Qy    537 gtttctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacag 596
          |||||||| | |||||||||||||||||| ||||||||||||||||||||||||||||||
Db   3061 GTTTCTGTTTTGTGTGTCTTTGTCTTGTGTGTCCTTGTCTACAGTTTTAATATGGGACAG 3120

Qy    597 acagtgactacccccttagtttgactctcgaccattggactgaagttagatccagggct 656
          || ||||| ||||| |||||||||||||||||||||||||||||||||||| ||||||||||
Db   3121 ACGGTGACGACCCCTCTTAGTTTGACTCTCGACCATTGGACTGAAGTTAAATCCAGGGCT 3180

Qy    657 cataatttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatgg 716
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   3181 CATAATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGTCTCTGAATGG 3240

Qy    717 ccaacattcgatgttggatggccatcagaggggacctttaattctgaaattatcctggct 776
          || |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   3241 CCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCTGAGATTATCCTGGCT 3300

Qy    777 gttaaggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatc 836
          ||||| |||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   3301 GTTAAAGCAATTATTTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAGCCCTATATC 3360

Qy    837 cttacgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaataaa 896
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||| |||||
Db   3361 CTTACGTGGCAAGATTTGGCAGAGGATCCTCCGCCATGGGTTAAACCATGGCTGAATAAG 3420

Qy    897 ccaagaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaa 956
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db   3421 CCAAGAAAGCCAGGTCCCCGAATTCTGGCTCTTGGAGAGAAAAACAAACACTCGGCTGAA 3480

Qy    957 aaagtcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccgga 1016
          |||||| ||||||| ||||| |||||||||||||| ||||| || ||| ||||||||||
Db   3481 AAAGTCAAGCCCTC-TCCTCATATCTACCCCGAGATTGAGGAACCACCGGCTTGGCCGGA 3539

Qy    1017 accccaacctgttccccccacccccttatccagcacagggtgctgtgagggga-cctctgc 1075
           ||||||| |||||||||||||||||||  |||||||||| | ||||||| ||| |||
Db   3540 ACCCCAATCTGTTCCCCCACCCCCTTATCTGGCACAGGGTGCCGCGAGGGGACCCTTTGC 3599

Qy    1076 ccctcctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgc 1135
           |||||||||||||||| |||||||||||| |||| ||||||||||||||||| ||||||
Db   3600 CCCTCCTGGAGCTCCGGCGGTGGAGGGACCTTCTGCAGGGACTCGGAGCCGGAGGGGCGC 3659

Qy    1136 caccccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccat 1195
           |||||||||||||||||||||||||||||||| |||||||||||||| || ||||||||||
Db   3660 CACCCCGGAGCGGACAGACGAGATCGCGACATTACCGCTGCGCACGTACGGCCCTCCCAC 3719

Qy    1196 gccaggggggccaattgcagcccctccagtattggccctttttcttctgcagatctctataa 1255
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3720 ACCGGGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAA 3779

Qy    1256 ttggaaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtgga 1315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3780 TTGGAAAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGGGGTTGGTGGA 3839

Qy    1316 gtcccttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacact 1375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3840 GTCCCTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACT 3899

Qy    1376 cttcacaaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggc 1435
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   3900 CTTCACAACCGAGGAGCGAGAGAGAATTCTATTAGAGGCTAGAAAAAATGTTCCTGGGGC 3959

Qy    1436 cgacgggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccc 1495
           |||||||||||||||| ||||||||||||||||||||||||||||||||| ||||||||
Db   3960 CGACGGGCGACCCACGCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTAACTCGCCC 4019

Qy    1496 cggttgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctct 1555
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4020 CGGTTGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGAAAATCTATCGCCAGGCTCT 4079

Qy    1556 ggtggcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggt 1615
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   4080 GGTGGCGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGGTAAGAGAAGT 4139

Qy    1616 gatgcagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcag 1675
           |||||||||||||| ||||| |||| || ||||||||||||| ||||||||||||||
Db   4140 GATGCAGGGACCGAATGAACCCCCCTCTGTTTTTCTTGAGAGGCTCTTGGAAGCCTTCAG 4199

Qy    1676 gcggttcacccctttttgatcctacctcagaggcccagaaagcctcagtggccctggcctt 1735
           ||||| |||||||||||||| |||||||||||||| |||||||||||||  |||||||
```

```
Db   4200 GCGGTACACCCCTTTTGATCCCACCTCAGAGGCCCAAAAAGCCTCAGTGGCTTTGGCCTT 4259

Qy   1736 cattgggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacagga 1795
          || || ||||| ||  ||||||| || ||||| ||||||||||||||||||||||||||||
Db   4260 TATAGGACAGTCAGCCTTGGATATTAGAAAGAAGCTTCAGAGACTGGAAGGGTTACAGGA 4319

Qy   1796 ggctgagttacgtgatctagtgagagaggcagagaaggtgttacagaagggagacaga 1855
          |||||||||||||||||||||||||  |||||||||| || ||||||| ||||||||||||
Db   4320 GGCTGAGTTACGTGATCTAGTGAAGGAGGCAGAGAAAGTATATTACAAAAGGGAGACAGA 4379

Qy   1856 agaggagaaggaacagagaaaagaaaaggagagagaagaaagggaggaaaagacgtgatag 1915
          ||| || | |||||| |||||||| |  || ||||| ||||||||||||||||||| |||
Db   4380 AGAAGAAAGGGAACAAAGAAAAGAGAGAGAAAGAGAGGAAAGGGAGGAAAGACGTAATAA 4439

Qy   1916 acggcaagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcag 1975
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db   4440 ACGGCAAGAGAAGAATTTGACTAAGATCTTGGCTGCAGTGGTTGAAGGGAAAAGCAATAC 4499

Qy   1976 ggagagagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctggg 2035
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4500 GGAAAGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGG 4559

Qy   2036 caataggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggc 2095
          |||||||||||||||||||||||||||| ||||| |||||||||||||| ||||||||||||
Db   4560 CAATAGGACCCCACTCGACAAGGACCAATGTGCATATTGTAAAGAAAGAGGACACTGGGC 4619

Qy   2096 aaggaactgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaagataa 2155
          |||||||||||||||||||||||||||||||| | | ||||||||||||||||||||||||
Db   4620 AAGGAACTGCCCCAAGAAGGGAAACAAAGGACCAAGGATCCTAGCTCTAGAAGAAGATAA 4679

Qy   2156 agattaggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtgga 2215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4680 AGATTAGGGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGTGGA 4739

Qy   2216 ggggcaaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaacc 2275
          |||||||||||||||||||||||||||||||||||| | |||||||||| ||||| ||
Db   4740 GGGGCAACCAGTTGAGTTCCTGGTTGATACCGGAGCGAAACATTCAGTGCTACTACAGCC 4799

Qy   2276 attaggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcagta 2335
          |||||||||||||||||| |||||||||||||||||||||| |||||||||||  |||||
Db   4800 ATTAGGAAAACTAAAAGATAAAAAATCCTGGGTGATGGGTG-CACAGGGCAACAACAGTA 4858

Qy   2336 tccatggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtttct 2395
          |||||||||||||||||| || ||||||||||||||||||||||||||||||||||||||
Db   4859 TCCATGGACTACCCGAAG-ACAGTTGACTTGGGAGTGGGACGGGTAACCCACTCGTTTCT 4917

Qy   2396 ggtcatccctgagtgcccagtacccccttctaggtagagacttactgaccaagatgggagc 2455
          |||||| |||||||||||||| ||||||  |||||||||||||| ||||||||||||||||
Db   4918 GGTCATACCTGAGTGCCCAGCACCCCTCTTAGGTAGAGACTTATTGACCAAGATGGGAGC 4977

Qy   2456 tcaaatttcttttgaacaaggaagaccagaagtgtctgtgaataacaaacccatcactgt 2515
          |||||||||||||||||| | ||||||||||||||  |||||||||| ||||||||||||
Db   4978 ACAAATTTCTTTTGAACAAGGGAAACCAGAAGTGTCTGCAAATAACAAACCTATCACTGT 5037

Qy   2516 gttgaccctccaattagatgatgaatatcgactatattctccccaagtaaagcctgatca 2575
          |||||||||||||||||||| ||||||||||||||| ||||||| ||||||||||||||||
Db   5038 GTTGACCCTCCAATTAGATGACGAATATCGACTATACTCTCCCCTAGTAAAGCCTGATCA 5097

Qy   2576 agatatacagtcctggttggagcagtttccccaagcctgggcagaaaccgcagggatggg 2635
          | ||||||| | |||||||| ||||||||||||||||||||||||||||||||||||||||
Db   5098 AAATATACAATTCTGGTTGGAACAGTTTCCCCAAGCCTGGGCAGAAACCGCAGGGATGGG 5157

Qy   2636 tttggcaaagcaagttcccccacaggttattcaactgaaggccagtgctacaccagtatc 2695
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   5158 TTTGGCAAAGCAAGTTCCCCCACAAGTTATTCAACTGAAGGCCAGTGCCACACCAGTGTC 5217

Qy   2696 agtcagacagtacccccttgagtagagaggctcgagaaggaatttggccgcatgttcaaag 2755
          |||||||||||||||||||||| ||| |||| |||||||||| |||||||||| ||||||
Db   5218 AGTCAGACAGTACCCCTTGAGTAAAGAAGCTCAAGAAGGAATTCGGCCGCATGTCCAAAG 5277

Qy   2756 attaatccaacagggcatcctagttcctgtccaatccccttggaatactcccctgctacc 2815
          |||||||||||||||||||||||||||||||||||| || ||||||||||||||||||||
Db   5278 ATTAATCCAACAGGGCATCCTAGTTCCTGTCCAATCTCCCTGGAATACTCCCCTGCTACC 5337

Qy   2816 ggttaggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtcaataa 2875
          |||||| |||||||||||| ||||| ||||||||||||||||||||||||||||||||||
Db   5338 GGTTAGAAAGCCTGGGACTAATGACTATCGACCAGTACAGGACTTGAGAGAGGTCAATAA 5397

Qy   2876 aagggtgcaggacatacacccaacggtcccgaacccttataacctcttgagcgccctccc 2935
          | |||||||||| |||||||||| |||||||||||||||||||||||| | || |||||
Db   5398 ACGGGTGCAGGATATACACCCAACAGTCCCGAACCCTTATAACCTCTTGTGTGCTCTCCC 5457

Qy   2936 gcctgaacggaactggtacacagtattggacttaaaagatgccttcttctgcctgagatt 2995
          ||  |||||| |||||| ||||||||||||||||| ||||||||||||||||||||||||
Db   5458 ACCCCAACGGAGCTGGTATACAGTATTGGACTTAAAGGATGCCTTCTTCTGCCTGAGATT 5517

Qy   2996 acaccccactagccaaccacttttttgccttcgaatggagagatccaggtacgggaagaac 3055
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5518 ACACCCCACTAGCCAACCACTTTTTGCCTTCGAATGGAGAGATCCAGGTACGGGAAGAAC 5577

Qy   3056 cgggcagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttga 3115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5578 CGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGA 5637

Qy   3116 cgaagccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcct 3175
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   5638 CGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATCCAACACCCTCAGGTGACCCTCCT 5697

Qy   3176 ccagtacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtac 3235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   5698 CCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACCAAACAGGACTGCTTAGAAGGCAC 5757

Qy   3236 gaaggcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggccca 3295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5758 GAAGGCACTACTGCTGGAATTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCA 5817

Qy   3296 gatttgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggct 3355
          |||||||||||||||||||||||||||||||||||||||||||||||| |||| ||||||
Db   5818 GATTTGCAGGAGAGAGGTAACATACTTGGGGTACAGTTTACGGGACGGGCAGCGATGGCT 5877

Qy   3356 gacggaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagt 3415
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   5878 GACGGAGGCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACAAAT 5937

Qy   3416 gagagagtttttggggacagctggattttgcagactgtggatcccggggtttgcgacctt 3475
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5938 GAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCGACCTT 5997

Qy   3476 agcagccccactctacccgctaaccaaagaaaaaggggggattctcctgggctcctgagca 3535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5998 AGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGGGAATTCTCCTGGGCTCCTGAGCA 6057

Qy   3536 ccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctccc 3595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6058 CCAGAAGGCATTTGATGCTATCAAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCC 6117

Qy   3596 tgacgtaactaaaacccttaccctttatgtggatgagcgtaagggagtagcccgaggagt 3655
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   6118 TGACGTAACTAAACCCTTTACCCTTTATGTGGATGAGCGTAAGGGAGTAGCCCGGGGAGT 6177

Qy   3656 tttaacccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagcttga 3715
          |||||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||
Db   6178 TTTAACCCAAACCCTAGGACCATGGAGAAGACCTGTCGCCTACCTGTCAAAGAAGCTCGA 6237

Qy   3716 tcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccatactggt 3775
          |||||||||||||||||||| |||| ||||||||||||||||||||||||||||||||||
Db   6238 TCCTGTAGCCAGTGGTTGGCCCATATGCCTGAAGGCTATCGCAGCTGTGGCCATACTGGT 6297

Qy   3776 caaggacgctgacaaattgactttgggac-agaatataactgtaatagcccccccatgcat 3834
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   6298 CAAGGACGCTGACAAATTGACTTTGGGACAAGAATATAACTGTAATAGCCCCCCATGCAT 6357

Qy   3835 tggagaacatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacccact 3894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6358 TGGAGAACATCGTTCGGCAGCCCCCAGACCGATGGATGACCAACGCCCGCATGACCCACT 6417
```

```
Qy   3895 atcaaagcctgcttctcacagagagggtcactttcgctccaccagccgctctcaaccctg 3954
          ||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||||
Db   6418 ATCAAAGCCTGCTTCTCACAGAGAGGGTCACGTTCGCTCCACCAACCGCTCTCAACCCTG 6477

Qy   3955 ccactcttctgcctgaagagactgatgaaccagtgactcatgattgccatcaactattga 4014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6478 CCACTCTTCTGCCTGAAGAGACTGATGAACCAGTGACTCATGATTGCCATCAACTATTGA 6537

Qy   4015 ttgaggagactggggtccgcaaggaccttacagacataccgctgactggagaagtgctaa 4074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6538 TTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGGAGAAGTGCTAA 6597

Qy   4075 cctggttcactgacggaagcagctatgtggtggaaggtaagaggatggctggggcggcag 4134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   6598 CCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGGTAAGAGGATGGCTGGGGCGGCGG 6657

Qy   4135 tggtggacgggacccgcacgatctgggccagcagcctgccggaaggaacttcagcgcaaa 4194
          |||||||||||||||||||||||||||||||||||||||| |||||||||||| ||||
Db   6658 TGGTGGACGGGACCCGCACGATCTGGGCCAGCAGCCTGCCGGGAGGAACTTCAGCACAAA 6717

Qy   4195 aggctgagctcatggccctcacgcaagctttgcggctggccgaagggaaatccataaaca 4254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6718 AGGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAAACA 6777

Qy   4255 tttatacggacagcaggtatgcctttgcgactgcacacgtacacggggccatctataaac 4314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6778 TTTATACGGACAGCAGGTATGCCTTTGCGACTGCACACGTACATGGGGCCATCTATAAAC 6837

Qy   4315 aaaggggttgcttacctcagcagggagggaaataaagaacaaagaggaaattctaagcc 4374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6838 AAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAGGAAATTCTAAGCC 6897

Qy   4375 tattagaagccttacatttgccaaaaaggctagctattatacactgtcctggacatcaga 4434
          |||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||
Db   6898 TATTAGAAGCCGTACATTTACCAAAAAGGCTAGCTATTATACACTGTCCTGGACATCAGA 6957

Qy   4435 aagccaaagatctcatatctagagggaaccagatggctgaccgggttgccaagcaggcag 4494
          |||| |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||
Db   6958 AAGCTAAAGATCTCATATCCAGAGGAAACCAGATGGCTGACCGGGTTGCCAAGCAGGCAG 7017

Qy   4495 cccaggctgttaaccttctgcctataatagaaacgcccaaagccccagaacccagacgac 4554
          |||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   7018 CCCAGGGTGTTAACCTTCTGCCTATAATAGAAATGCCCAAAGCCCCAGAACCCAGACGAC 7077

Qy   4555 agtacaccctagaagactggcaagagataaaaaagatagaccagttctctgagactccgg 4614
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   7078 AGTACACCCTAGAAGACTGGCAAGAGATAAAAAAGATAGACCA-TTCTCTGAGACTCCGG 7136

Qy   4615 aggggacctgctatacctcatatgggaaggaaatcctgccccacaaagaagggttagaat 4674
          | ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   7137 AAGGGACCTGCTATACCTCAGATGGGAAGGAAATCCTGCCCCACAAAGAAGGGTTAGAAT 7196

Qy   4675 atgtccaac-agatacatcgtctaacccacctaggaactaaacacctgcagcagttggtc 4733
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7197 ATGTCCAACAAGATACATCGTCTAACCCACCTAGGAACTAAACACCTGCAGCAGTTGGTC 7256

Qy   4734 agaacatccccttatcatgttctgaggctaccaggagtggctgactcggtggtcaaacat 4793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7257 AGAACATCCCCTTATCATGTTCTGAGGCTACCAGGAGTGGCTGACTCGGTGGTCAAACAT 7316

Qy   4794 tgtgtgccctgccagctggttaatgctaatccttccagaatacctccaggaaagagacta 4853
          |||||||||||||||||||||||||||||||||||||||| ||||||| |||||||||||
Db   7317 TGTGTGCCCTGCCAGCTGGTTAATGCTAATCCTTCCAGAATGCCTCCAGGGAAGAGACTA 7376

Qy   4854 aggggaagccacccaggcgctcactgggaagtggacttcactgaggtaaagccggctaaa 4913
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7377 AGGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAA 7436

Qy   4914 tacggaaacaaatatctattggtttttgtagacacctttcaggatgggtagaggcttat 4973
          || |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   7437 TATGGAAACAAATACCTATTGGTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTAT 7496

Qy   4974 cctactaagaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttcca 5033
          |||||||||||||||||||||||||||||||| ||||| |||||||||| ||||||||||||
Db   7497 CCTACTAAGAAAGAGACTTCAACCGTGGTAGCTAAAAAAATACTGGAAGAAATTTTTCCA 7556

Qy   5034 agatttggaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccaggta 5093
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   7557 AGATTTGGAATACCTAAGGTAATAGGGTCAGACAATGGTCCAGCTTTTGTTGCCCAGGTA 7616

Qy   5094 agtcagggactggccaagatattggggattgattggaaactgcattgtgcatacagaccc 5153
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7617 AGTCAGGGACTGGCCAAGATATTGGGGATTGATTGGAAACTGCATTGTGCATACAGACCC 7676

Qy   5154 caaagctcaggacaggtagagaggatgaatagaaccattaaagagacccttaccaaattg 5213
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   7677 CAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACCCTTACTAAATTG 7736

Qy   5214 accacagagactggcattaatgattggatggctctcctgccctttgtgctttttagggtg 5273
          ||| | |||||||||| |||||||||||||| ||||||||||||||||||||||||||||
Db   7737 ACCGCGGAGACTGGCGTTAATGATTGGATAGCTCTCCTGCCCTTTGTGCTTTTTAGGGTT 7796

Qy   5274 aggaacacccctggacagtttgggctgacccccctatgaattgctctacgggggaccccc 5333
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   7797 AGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTACTCTACGGGGGACCCCCC 7856

Qy   5334 ccgttggcagaaattgcctttgcacatagtgctgatgtgctgctttcccagcctttgttc 5393
          || |||| |||||||| | || ||||||||||||||||||||||||||||||||||||||
Db   7857 CCATTGGTAGAAATTGCTTCTGTACATAGTGCTGATGTGCTGCTTTCCCAGCCTTTGTTC 7916

Qy   5394 tctaggctcaaggcgctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcc 5453
          |||||||||||||| || |||||||||| || |||||||||| ||| ||||||||||||
Db   7917 TCTAGGCTCAAGGCACTTGAGTGGGTGAGACAACGAGCGTGGAGGCAACTCCGGGAGGCC 7976

Qy   5454 tactc---aggaggagacttgcaagttccacatcgcttccaagttggagattcagtctat 5510
          |||||   |||||||||||||||  | |||||||| |||||||| ||||||||||||||
Db   7977 TACTCAGGAGGAGGAGACTTGCAGATCCCACATCGTTTCCAAGTGGGAGATTCAGTCTAC 8036

Qy   5511 gttagacgccaccgtgcaggaaac 5534
          |||||||||||||||||||||||||
Db   8037 GTTAGACGCCACCGTGCAGGAAAC 8060
```

RESULT 4
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

Query Match 41.1%; Score 3376.6; DB 3; Length 3482;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 3441; Conservative 0; Mismatches 29; Indels 7; Gaps 4;

```
Qy   4718 cctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctga 4777
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 cctgcagcagttggtcagaacatccccttatcatgttctgaggctaccaggagtggctga 65

Qy   4778 ctcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacc 4837
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

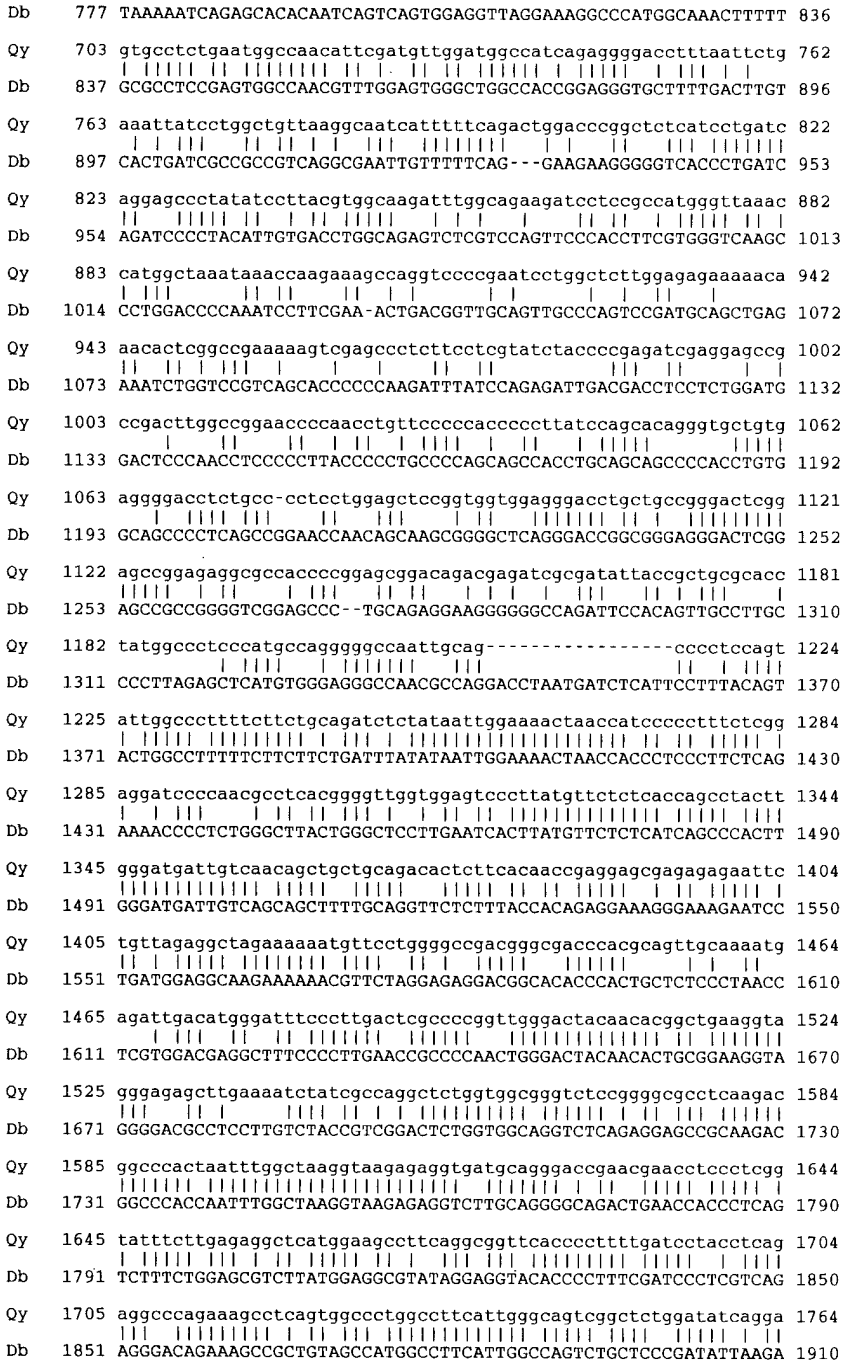
```
Db      66 ctcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttccagaatacc 125

Qy    4838 tccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactga 4897
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126 tccaggaaagagactaaggggaagccacccaggcgctcactgggaagtggacttcactga 185

Qy    4898 ggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcagg 4957
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     186 ggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacacctttcagg 245

Qy    4958 atgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaagaaaatact 5017
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     246 atgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaaaaaaatact 305

Qy    5018 ggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaatggtccagc 5077
           |||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db     306 ggaggaaatttttccgagatttggaatacctaaggtaatcgggtcagacaatggtccagc 365

Qy    5078 tttcgttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgca 5137
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     366 ttttgttgcccaggtaagtcagggactggccaagatattggggattgattggaaactgca 425

Qy    5138 ttgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaaga 5197
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     426 ttgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaaccattaaaga 485

Qy    5198 gacccttaccaaattgaccacagagactggcattaatgattggatggctctcctgccctt 5257
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     486 gacccttaccaaattgaccacagagactggcattaatgattggatagctctcctgccctt 545

Qy    5258 tgtgctttttagggtgaggaacacccctggacagtttgggctgacccccatatgaattgct 5317
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     546 tgtgctttttagggttaggaacacccctggacagtttgggctgacccccatatgaattgct 605

Qy    5318 ctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctgatgtgctgct 5377
           |||||||||||||||||||||| ||||||||| | || ||||||||||||||||||||||
Db     606 ctacgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgct 665

Qy    5378 ttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcgagcgtggaa 5437
           ||||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     666 ttcccagcctctgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaa 725

Qy    5438 gcagctccgggaggcctactcagg---aggagacttgcaagttccacatcgcttccaagt 5494
           |||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db     726 gcagctccgggaggcctactcaggagaaggagacttgcaagttccacatcgcttccaagt 785

Qy    5495 tggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaaggg 5554
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     786 gggagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaaggg 845

Qy    5555 accttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatccccacctg 5614
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     846 cccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctg 905

Qy    5615 gatccatgcatcccacgttaagccggcgccacctcccgactcggggtggagagccgaaaa 5674
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     906 gatccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaa 965

Qy    5675 gactgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactc 5734
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     966 gactgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactc 1025

Qy    5735 cccaggccagtagtaaacgccttatagacagctcgaaccccccatagacctttatccctta 5794
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1026 cccaggccagtagtaaacgccttatagacagctcgaaccccccatagacctttatccctta 1085

Qy    5795 cctggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctc 5854
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1086 cctggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctc 1145

Qy    5855 ctagaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgtta 5914
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1146 ctagaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgtta 1205

Qy    5915 aaagcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagaga 5974
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1206 aaagcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagaga 1265

Qy    5975 aagagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacct 6034
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1266 aagagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacct 1325

Qy    6035 ccaacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttg 6094
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1326 ccaacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttg 1385

Qy    6095 tcaattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctccc 6154
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    1386 tcaattccggcccgggcaagtacaaagtgatgaaactatataaagataagagctgctccc 1445

Qy    6155 catcagacttagattatctaaagataagtttcactg--aaaggaaaacaggaaaatattc 6212
           |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db    1446 catcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattc 1505

Qy    6213 aaaagtggataaatggtatgagctggggaatagttttttat-tatatggcgggggagcag 6271
           |||||||||||||||||||||||||||||||||||||||||  |||||||||||||||||
Db    1506 aaaagtggataaatggtatgagctggggaatagttttttataaatatggcgggggagcag 1565

Qy    6272 ggtccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatgg 6331
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    1566 ggtccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcagtgg 1625

Qy    6332 gacccgataaagtactggctgaacaggggcccccggccctggagccaccgcataacttgc 6391
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1626 gacccgataaagtactggctgaacaggggcccccggccctggagccaccgcataacttgc 1685

Qy    6392 cggtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtacca 6451
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    1686 cggtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacggtacca 1745

Qy    6452 ctggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggac 6511
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1746 ctggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggac 1805

Qy    6512 agagactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatg 6571
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1806 agagactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatg 1865

Qy    6572 ccacttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggcta 6631
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1866 ccacttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggcta 1925

Qy    6632 aagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaa 6691
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1926 aagaaggaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaa 1985

Qy    6692 ataagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccat 6751
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1986 ataagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccat 2045

Qy    6752 cccaccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtatt 6811
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2046 cccaccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtatt 2105

Qy    6812 tagtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttcca 6871
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2106 tagtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttcca 2165

Qy    6872 cctcagtcttcaaccaatccaaagatttctgtgtcatggtccaaatcgtcccccgagtgt 6931
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2166 cctcagtcttcaaccaatccaaagatttctgtgtcatggtccaaatcgtcccccgagtgt 2225

Qy    6932 actaccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaa 6991
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2226 actaccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaa 2285
```

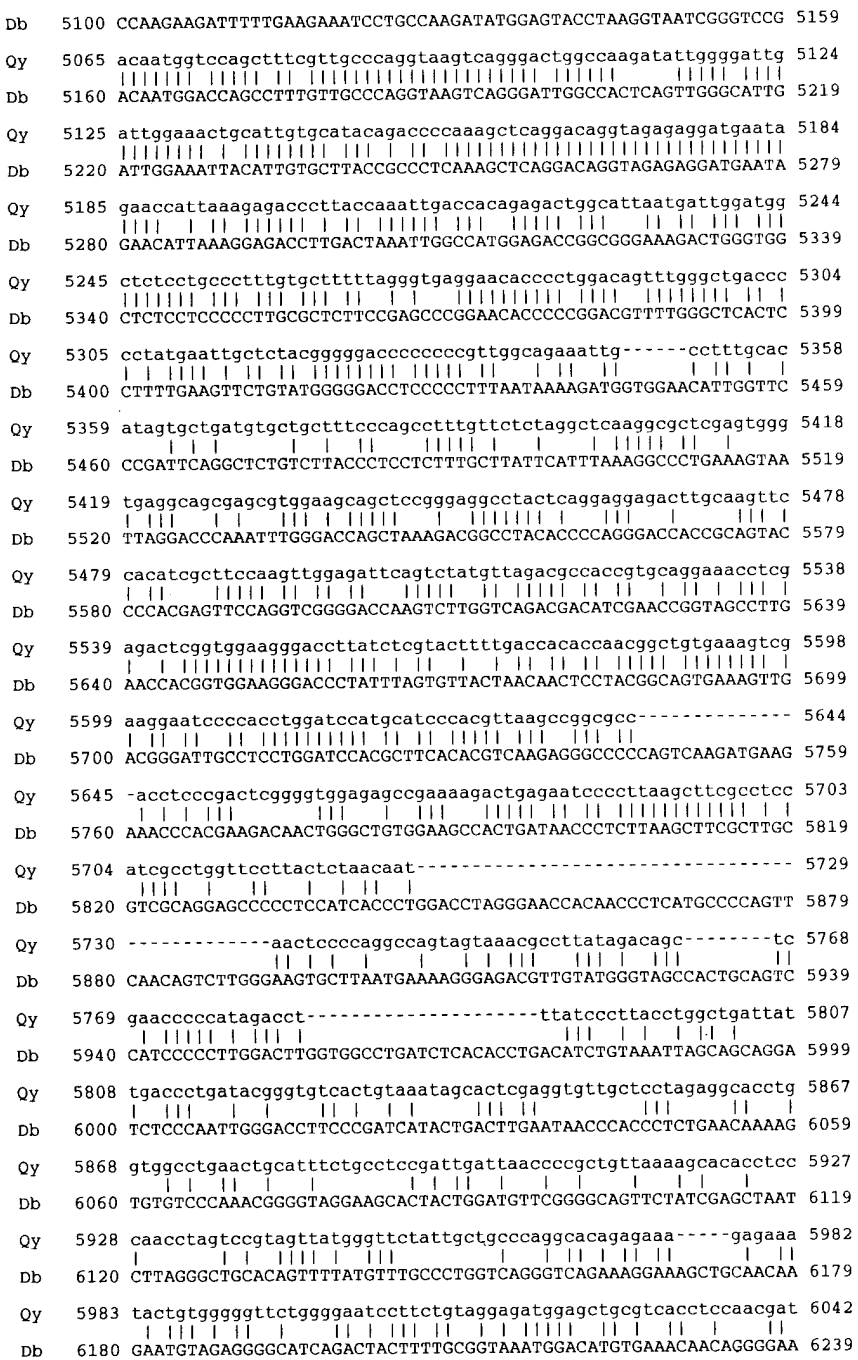
```
Qy   6992 gagaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtag 7051
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2286 gagaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtag 2345

Qy   7052 gaacagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagc 7111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2346 gaacagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagc 2405

Qy   7112 tacatgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaag 7171
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   2406 tacatgcggccatgacagaagatctccgagccttagaggagtctgttagcaacctagaag 2465

Qy   7172 agtccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgt 7231
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2466 agtccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgt 2525

Qy   7232 ttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatc 7291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2526 ttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatc 2585

Qy   7292 actcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaa 7351
          |||||||||||||||||||||||||| |||||||||| ||| ||||||||||||||||||||||
Db   2586 actcaggagccatcagagactccatgagcaagcttagagaaaggttagagaggcgtcgaa 2645

Qy   7352 gggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatga 7411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2646 gggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatga 2705

Qy   7412 ccaccctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttg 7471
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   2706 ccaccctgctttctgctctgacgggacccctagtagtcctgctcctgttacttacagttg 2765

Qy   7472 ggccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccaga 7531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2766 ggccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccaga 2825

Qy   7532 tcatggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctagc 7591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2826 tcatggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctagc 2885

Qy   7592 cttcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaa 7651
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2886 cttcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaa 2945

Qy   7652 aatgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgcccccgaa 7711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2946 aatgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgcccccgaa 3005

Qy   7712 ttacagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagg 7771
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3006 ttccagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagg 3065

Qy   7772 gcctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctgga 7831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3066 gcctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctgga 3125

Qy   7832 ttctgtaaaactgactggcaccatagaagaattgattacacattgacagccctagtgacc 7891
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3126 ttctgtaaaaccgactggcaccatagaagaattgattacacattgacagccctagtgacc 3185

Qy   7892 tatctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagc 7951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3186 tatctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagc 3245

Qy   7952 tattttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattggtcca 8011
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3246 tattttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattggtcca 3305

Qy   8012 tggagcgcgggctc-tcgatattttaaaatgattggtttgtgacgcacaggctttgttgt 8070
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3306 cggagcgcgggctcttcgatattttaaaatgattggtttgtgacgcacaggctttgttgt 3365

Qy   8071 gaaccccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcgtg 8130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3366 gaaccccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcgtg 3425

Qy   8131 gtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 8187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3426 gtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatc 3482
```

RESULT 5
US-09-075-272-1
; Sequence 1, Application US/09075272
; Patent No. 6136598
; GENERAL INFORMATION:
; APPLICANT: MILLER, A. DUSTY
; APPLICANT: WOLGAMOT, GREG
; APPLICANT: BONHAM, LYNN
; TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL
; TITLE OF INVENTION: PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/075,272
; FILING DATE: 08-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/046,140
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Poor, Brian W.
; REGISTRATION NUMBER: 32,928
; REFERENCE/DOCKET NUMBER: 14538A-003710
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8655 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-075-272-1

Query Match 28.0%; Score 2295.8; DB 3; Length 8655;
Best Local Similarity 60.3%; Pred. No. 0;
Matches 4362; Conservative 0; Mismatches 2662; Indels 209; Gaps 26;

```
Qy    464 ccctccgcggccgtccgactcttttgcctgcttgtggaagacgcggacgggtcgcgtgtg 523
          ||  || |  || || || |||||    ||  |||||  ||||||  |   | |||  ||
Db    601 CCGACCCCATCCATCTGAATCTTTAAGTTGTCTGTGGTCGACGCGAAGTCGCCGCCGGTT 660

Qy    524 tctggatc-tgttggtttctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagtt 582
          |  |  || | || || || |||||||||    |  | |||||  | ||  |    | ||
Db    661 TTGGTTTCTTTTTTGTCTCAGTCTCGTGTCCGCTCTTGTTGTGTCTGCTATT----ATTT 716

Qy    583 ttaatatgggacagacagtgactacccccctttagtttgactctcgaccattggactgaag 642
          |  | ||||||||  | ||  | || ||||||  | | || || || |||||||  || |
Db    717 TAGAAATGGGACAATCTGTCTCCACTCCCCTTTCTCTAACCCTGGAGCATTGGAAGGAGG 776

Qy    643 ttagatccagggctcataatttgtcagttcaggttaagaagggaccttggcagactttct 702
          | | |  ||| || || |||  ||||||  |||||| ||| || || ||||| ||||| |
```


Db 777 TAAAAATCAGAGCACACAATCAGTCAGTGGAGGTTAGGAAAGGCCCATGGCAAACTTTTT 836
Qy 703 gtgcctctgaatggccaacattcgatgttggatggccatcagaggggaccttaattctg 762
Db 837 GCGCCTCCGAGTGGCCAACGTTTGGAGTGGGCTGGCCACCGGAGGGTGCTTTTGACTTGT 896
Qy 763 aaattatcctggctgttaaggcaatcatttttcagactggacccggctctcatcctgatc 822
Db 897 CACTGATCGCCGCCGTCAGGCGAATTGTTTTTCAG---GAAGAAGGGGGTCACCCTGATC 953
Qy 823 aggagccctatatccttacgtggcaagatttggcagaagatcctccgccatgggttaaac 882
Db 954 AGATCCCCTACATTGTGACCTGGCAGAGTCTCGTCCAGTTCCCACCTTCGTGGGTCAAGC 1013
Qy 883 catggctaaataaaccaagaaagccaggtccccgaatcctggctcttggagagaaaaaca 942
Db 1014 CCTGGACCCCAAATCCTTCGAA-ACTGACGGTTGCAGTTGCCCAGTCCGATGCAGCTGAG 1072
Qy 943 aacactcggccgaaaaagtcgagccctcttcctcgtatctaccccgagatcgaggagccg 1002
Db 1073 AAATCTGGTCCGTCAGCACCCCCCAAGATTTATCCAGAGATTGACGACCTCCTCTGGATG 1132
Qy 1003 ccgacttggccggaaccccaacctgttcccccacccccttatccagcacagggtgctgtg 1062
Db 1133 GACTCCCAACCTCCCCCTTACCCCCTGCCCCAGCAGCCACCTGCAGCAGCCCCACCTGTG 1192
Qy 1063 aggggacctctgcc-cctcctggagctccggtggtggagggacctgctgccgggactcgg 1121
Db 1193 GCAGCCCCTCAGCCGGAACCAACAGCAAGCGGGGCTCAGGGACCGGCGGGAGGGACTCGG 1252
Qy 1122 agccggagaggcgccaccccggagcggacagacgagatcgcgatattaccgctgcgcacc 1181
Db 1253 AGCCGCCGGGGTCGGAGCCC--TGCAGAGGAAGGGGGGCCAGATTCCACAGTTGCCTTGC 1310
Qy 1182 tatggccctcccatgccaggggcccaattgcag----------cccctccagt 1224
Db 1311 CCCTTAGAGCTCATGTGGGAGGGCCAACGCCAGGACCTAATGATCTCATTCCTTTACAGT 1370
Qy 1225 attggccctttttcttctgcagatctctataattggaaaactaaccatccccctttctcgg 1284
Db 1371 ACTGGCCTTTTTCTTCTTCTGATTTATATAATTGGAAAACTAACCACCCTCCCTTCTCAG 1430
Qy 1285 aggatccccaacgcctcacgggggttggtggagtcccttatgttctctcaccagcctactt 1344
Db 1431 AAAACCCCTCTGGGCTTACTGGGCTCCTTGAATCACTTATGTTCTCTCATCAGCCCACTT 1490
Qy 1345 gggatgattgtcaacagctgctgcagacactcttcacaaccgaggagcgagagagaattc 1404
Db 1491 GGGATGATTGTCAGCAGCTTTTGCAGGTTCTCTTTACCACAGAGGAAAGGGAAAGAATCC 1550
Qy 1405 tgttagaggctagaaaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatg 1464
Db 1551 TGATGGAGGCAAGAAAAAACGTTCTAGGAGAGGACGGCACACCCACTGCTCTCCCTAACC 1610
Qy 1465 agattgacatgggatttcccttgactcgccccggttgggactacaacacggctgaaggta 1524
Db 1611 TCGTGGACGAGGCTTTCCCCTTGAACCGCCCCAACTGGGACTACAACACTGCGGAAGGTA 1670
Qy 1525 gggagagcttgaaaatctatcgccaggctctggtggcgggtctccggggcgcctcaagac 1584
Db 1671 GGGGACGCCTCCTTGTCTACCGTCGGACTCTGGTGGCAGGTCTCAGAGGAGCCGCAAGAC 1730
Qy 1585 ggcccactaatttggctaaggtaagagaggtgatgcagggaccgaacgaacctccctcgg 1644
Db 1731 GGCCCACCAATTTGGCTAAGGTAAGAGAGGTCTTGCAGGGGCAGACTGAACCACCCTCAG 1790
Qy 1645 tatttcttgagaggctcatggaagccttcaggcggttcacccctttgatcctacctcag 1704
Db 1791 TCTTTCTGGAGCGTCTTATGGAGGCGTATAGGAGGTACACCCCTTTCGATCCCTCGTCAG 1850
Qy 1705 aggcccagaaagcctcagtggccctggccttcattgggcagtcggctctggatatcagga 1764
Db 1851 AGGGACAGAAAGCCGCTGTAGCCATGGCCTTCATTGGCCAGTCTGCTCCCGATATTAAGA 1910
Qy 1765 agaaacttcagagactggaagggttacaggaggctgagttacgtgatctagtgagagagg 1824
Db 1911 AAAAGCTACAGAGGCTGGAGGGGCTCCAAGATTATACGCTCCAAGATTTAGTGAAGGAAG 1970
Qy 1825 cagagaaggtgtattacagaagggagacagaagaggagaaggaacagagaaaagaaaagg 1884
Db 1971 CAGAGAAGGTTTATCACAAGAGAGAAACAGAAGAAGAGAGGCAGGAGAGAGAAAAGAAAG 2030
Qy 1885 agagagaagaaagggaggaaagacgtgatagacggcaagagaagaatttgactaagatct 1944
Db 2031 AAGTAGAGGAGAGGGAAAATAGGCGAGATCGCCGTCAGGAGAGAAATTTGAGTAAGATTC 2090
Qy 1945 tggccgcagtggttgaagggaagagcagcagggagagagagagagattttaggaaaatta 2004
Db 2091 TGGCCGCAGTTATAAATG---------ATAGGCAGTCAGAAAAAGGCAGAACAGGATTCC 2141
Qy 2005 ggtcaggccctagacagtcagggaacctgggcaataggaccccactcgacaaggaccagt 2064
Db 2142 TGGGCAACAGGGCAGTGAAACCGCCAGGTGGCAGAAAGACGCCGCTGGAAAAAGACCAAT 2201
Qy 2065 gtgcgtattgtaaagaaaaaggacactgggcaaggaactgccccaagaagggaaacaaag 2124
Db 2202 GCGCCTTTTGCAAAGAAAAAGGACACTGGGCTAAAGACTGCCCTAAGAAAAGAAGGCA-- 2259
Qy 2125 gaccgaaggtcctagctctagaagaagataaagattaggggagacggggttcggacccccc 2184
Db 2260 -ATTCAAGGTCCTGACCCTAGAAG------ACGATTAGGGAAGTCGGGGCTCAGACCCCC 2312
Qy 2185 tccccgagcccagggtaactttgaaggtggaggggcaaccagttgagttcctggttgata 2244
Db 2313 TCCCTGAGCCTAGGGTAACTTTGTCTGTGGAGGGGACTCCCGTTAATTTCCTGATAGATA 2372
Qy 2245 ccggagcggagcattcagtgctgctacaaccattaggaaaactaaaagaaaaaaaatcct 2304
Db 2373 CTGGAGCAGAGCATTCGGTACTCACCAGCCCCCTGGGCAAACTAGGCTCTAAAAGGACCA 2432
Qy 2305 gggtgatgggtgccacagggcaacggcagtatccatggactacccgaagaaccgttgact 2364
Db 2433 TAGTGGTTGGAGCCACTGGGAGTAAACTTTACCCCTGGACAACCAAAAGAGCCTTACAGA 2492
Qy 2365 tgggagtgggacgggtaacccactcgtttctggtcatccctgagtgcccagtacccctttc 2424
Db 2493 TAGATAAAAACATGGTGACCCACTCTTTCCTGGTGATACCTGAGTGTCCTGCTCCCCTTC 2552
Qy 2425 taggtagagacttactgaccaagatgggagctcaaatttcttttgaacaaggaagaccag 2484
Db 2553 TGGGACGTGATCTGTTAACAAAACTAAAGGCTCAAGTTCAATTTACTTCAGAAGGTCCAC 2612
Qy 2485 aagtgtctgtgaataacaaacccatcactgtgttgaccctccaattagatgatgaatatc 2544
Db 2613 AAGTAAGCTGGGGAAAAGCCCCCCTTGCTTGCCTTGTCCTCAGCACGGAAGAAGAGTACC 2672
Qy 2545 gactatattctccccaagtaaagcctgatcaagatatacagtcctggttggagcagtttc 2604
Db 2673 GGTTGCATGAAGAGCAACCCAAAGGTGCAGCCCCTTTAGA---CTGGGTAACTGCGTTCC 2729
Qy 2605 cccaagcctgggcagaaaccgcagggatgggtttggcaaagcaagttcccccacaggtta 2664
Db 2730 CCAATGTCTGGGCGGAACAAGCAGGGATGGGGTTGGCTAAACAAGTGCCTCCAGTCGTGG 2789
Qy 2665 ttcaactgaaggccagtgctacaccagtatcagtcagacagtacccccttgagtagagagg 2724
Db 2790 TAGAACTTAAGGCTGATGCTACCCCCATCTCAGTAAGACAATACCCCATGAGCAAGGAAG 2849
Qy 2725 ctcgagaaggaatttggccgcatgttcaaagattaatccaacagggcatcctagttcctg 2784
Db 2850 CTAAGGAGGGCATCCGACCTCATATTCGGAGGCTGCTAGACCAAGGAATTTTAGTGGCCT 2909
Qy 2785 tccaatcccccttggaatactcccctgctaccggttaggaagcctgggaccaatgattatc 2844
Db 2910 GTCAGTCCCCCTGGAACACTCCACTTTTGCCAGTACGAAAGCCAGGGACCAATGATTATC 2969


Db 5100 CCAAGAAGATTTTTGAAGAAATCCTGCCAAGATATGGAGTACCTAAGGTAATCGGGTCCG 5159
Qy 5065 acaatggtccagctttcgttgcccaggtaagtcagggactggccaagatattggggattg 5124
Db 5160 ACAATGGACCAGCCTTTGTTGCCCAGGTAAGTCAGGGATTGGCCACTCAGTTGGGCATTG 5219
Qy 5125 attggaaactgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaata 5184
Db 5220 ATTGGAAATTACATTGTGCTTACCGCCCTCAAAGCTCAGGACAGGTAGAGAGGATGAATA 5279
Qy 5185 gaaccattaaagagaccctaccaaattgaccacagagactggcattaatgattggatgg 5244
Db 5280 GAACATTAAAGGAGACCTTGACTAAATTGGCCATGGAGACCGGCGGGAAAGACTGGGTGG 5339
Qy 5245 ctctcctgccctttgtgctttttagggtgaggaacacccctggacagtttgggctgaccc 5304
Db 5340 CTCTCCTCCCCCTTGCGCTCTTCCGAGCCCGGAACACCCCCGGACGTTTTGGGCTCACTC 5399
Qy 5305 cctatgaattgctctacgggggaccccccccgttggcagaaattg------cctttgcac 5358
Db 5400 CTTTTGAAGTTCTGTATGGGGGACCTCCCCCTTTAATAAAAGATGGTGGAACATTGGTTC 5459
Qy 5359 atagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcgagtggg 5418
Db 5460 CCGATTCAGGCTCTGTCTTACCCTCCTCTTTGCTTATTCATTTAAAGGCCCTGAAAGTAA 5519
Qy 5419 tgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagacttgcaagttc 5478
Db 5520 TTAGGACCCAAATTTGGGACCAGCTAAAGACGGCCTACACCCCAGGGACCACCGCAGTAC 5579
Qy 5479 cacatcgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaaacctcg 5538
Db 5580 CCCACGAGTTCCAGGTCGGGGACCAAGTCTTGGTCAGACGACATCGAACCGGTAGCCTTG 5639
Qy 5539 agactcggtggaagggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcg 5598
Db 5640 AACCACGGTGGAAGGGACCCTATTTAGTGTTACTAACAACTCCTACGGCAGTGAAAGTTG 5699
Qy 5599 aaggaatccccacctggatccatgcatcccacgttaagccggcgcc---------- 5644
Db 5700 ACGGGATTGCCTCCTGGATCCACGCTTCACACGTCAAGAGGGCCCCCAGTCAAGATGAAG 5759
Qy 5645 -acctcccgactcggggtggagagccgaaaagactgagaatcccccttaagcttcgcctcc 5703
Db 5760 AAACCCACGAAGACAACTGGGCTGTGGAAGCCACTGATAACCCTCTTAAGCTTCGCTTGC 5819
Qy 5704 atcgcctggttccttactctaacaat---------- 5729
Db 5820 GTCGCAGGAGCCCCCTCCATCACCCTGGACCTAGGGAACCACAACCCTCATGCCCCAGTT 5879
Qy 5730 ----------aactccccaggccagtagtaaacgccttatagacagc--------tc 5768
Db 5880 CAACAGTCTTGGGAAGTGCTTAATGAAAAGGGAGACGTTGTATGGGTAGCCACTGCAGTC 5939
Qy 5769 gaaccccccatagacct----------ttatcccttacctggctgattat 5807
Db 5940 CATCCCCCTTGGACTTGGTGGCCTGATCTCACACCTGACATCTGTAAATTAGCAGCAGGA 5999
Qy 5808 tgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctg 5867
Db 6000 TCTCCCAATTGGGACCTTCCCGATCATACTGACTTGAATAACCCACCCTCTGAACAAAAG 6059
Qy 5868 gtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcc 5927
Db 6060 TGTGTCCCAAACGGGGTAGGAAGCACTACTGGATGTTCGGGGCAGTTCTATCGAGCTAAT 6119
Qy 5928 caacctagtccgtagttatgggttctattgctgcccaggcacagagaaa-----gagaaa 5982
Db 6120 CTTAGGGCTGCACAGTTTTATGTTTGCCCTGGTCAGGGTCAGAAAGGAAAGCTGCAACAA 6179
Qy 5983 tactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacctccaacgat 6042
Db 6180 GAATGTAGAGGGGCATCAGACTACTTTTGCGGTAAATGGACATGTGAAACAACAGGGGAA 6239
Qy 6043 ggagactggaa----atggccgatctctctccaggaccgggtaaaattctcctttgtcaa 6098
Db 6240 GCCTACTGGAAGCCTTCCGCTGATTGGGACCTGATCACGGTAAAACGTGGTAGTGGTTAT 6299
Qy 6099 ttccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctccccatc 6158
Db 6300 GATAAGCCAAACCAAGGAGAAAGAAACCCATACAAATACCTAGATTCTGGGTGTGCTCTT 6359
Qy 6159 agacttagattatctaaagataagtttcactgaaaggaaaacaggaaaatattc--aaaa 6216
Db 6360 AAAAATTACAGCCCCCCAGGACCATGCAAAGGTAAATACTGCAACCCCCTACTCATAAAA 6419
Qy 6217 gtggataaatggtatgagctggggaatagttttttattatatggcgggggagcagggtcc 6276
Db 6420 TTCACTGAGAAAGGGAAACAAGCTCGCCTGAGTTGGCTTAAAGGAAATAGGTGGGGTTGG 6479
Qy 6277 actttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatgggaccc 6336
Db 6480 CGAGTATATATTCCAATAAGAGACCCTGGGTTTATCTTTACGATTAGACTGACAGTAAGA 6539
Qy 6337 gataaagtactggctgaacaggggccccggccctggagccaccgcataacttgccggtg 6396
Db 6540 GACCTGGCAGTAACATCCATAGGACCCAACAAGGTCCTTACGGAACAGGCCCCCCAGTT 6599
Qy 6397 ccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccactgga 6456
Db 6600 GCACCGGCTCCCCCGAGAGTCCCAGCCGTGCCAGCTCCACCAACTTCACGGCCCTACACA 6659
Qy 6457 ttgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacagaga 6516
Db 6660 GTAGGACCCTCATTAGAGACAACCTTAGCCTCCCCACCACTCCTAGATACAGAAAACCGT 6719
Qy 6517 ctcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgccact 6576
Db 6720 CTGGTCAGTCTAGTTCAGGGAGCCTTTTTAGTTTTAAATAGGACTAATCCTAATATGACT 6779
Qy 6577 tcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaagaa 6636
Db 6780 CAATCATGTTGGTTATGCTATGCCTCTAACCCCCCTTATTATGAAGGAATTGCTCAGACT 6839
Qy 6637 agaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaataag 6696
Db 6840 AGAACTTACAATATTACTTCAGATCATTC---TCAATGTCTTTGGGGAGAGAACAGGAAG 6896
Qy 6697 cttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcccac 6756
Db 6897 TTGACTCTGACAGCAGTTTCAGGAAATGGGCTTTGTTTAGGTCAGGTGCCCCAGGATAAA 6956
Qy 6757 caacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtatttagta 6816
Db 6957 TGGCACCTCTGTAACCAGACTCAAAATATCCGACCTAACAAAGGTGGTCAGTATCTAGTG 7016
Qy 6817 cctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacctca 6876
Db 7017 CCTCCCATAGACACAGTATGGGCTTGCAATACAGGTCTCACTCCTTGTATATCTATGTCT 7076
Qy 6877 gtcttcaaccaatccaaagatttctgtgtcatggtccaaatcgtcccccgagtgtactac 6936
Db 7077 GTTTTCAATAGCTCCAAAGATTTCTGTATTTTAGTTCAGCTTATTCCTAGACTCCTGTAT 7136
Qy 6937 catcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaagagaa 6996
Db 7137 CATGATGATAGCTCATTTCTAGACAAATTTGAACATCGG---GTCCGCTGGAAAAGAGAA 7193
Qy 6997 cccgtatcccttaccctagctgtaatgctcggatt---agggacggccgttggcgtagga 7053
Db 7194 CCCATTACTTTAACGTTAGCAGTGCTTTTAGGATTGGGAGTGGCCGCAGCTGGGGTAGGT 7253
Qy 7054 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 7113
Db 7254 ACTGGAACCGCTGCCTTAATCCAGACCCCCCGATACTTTGAG----------GAATTA 7301

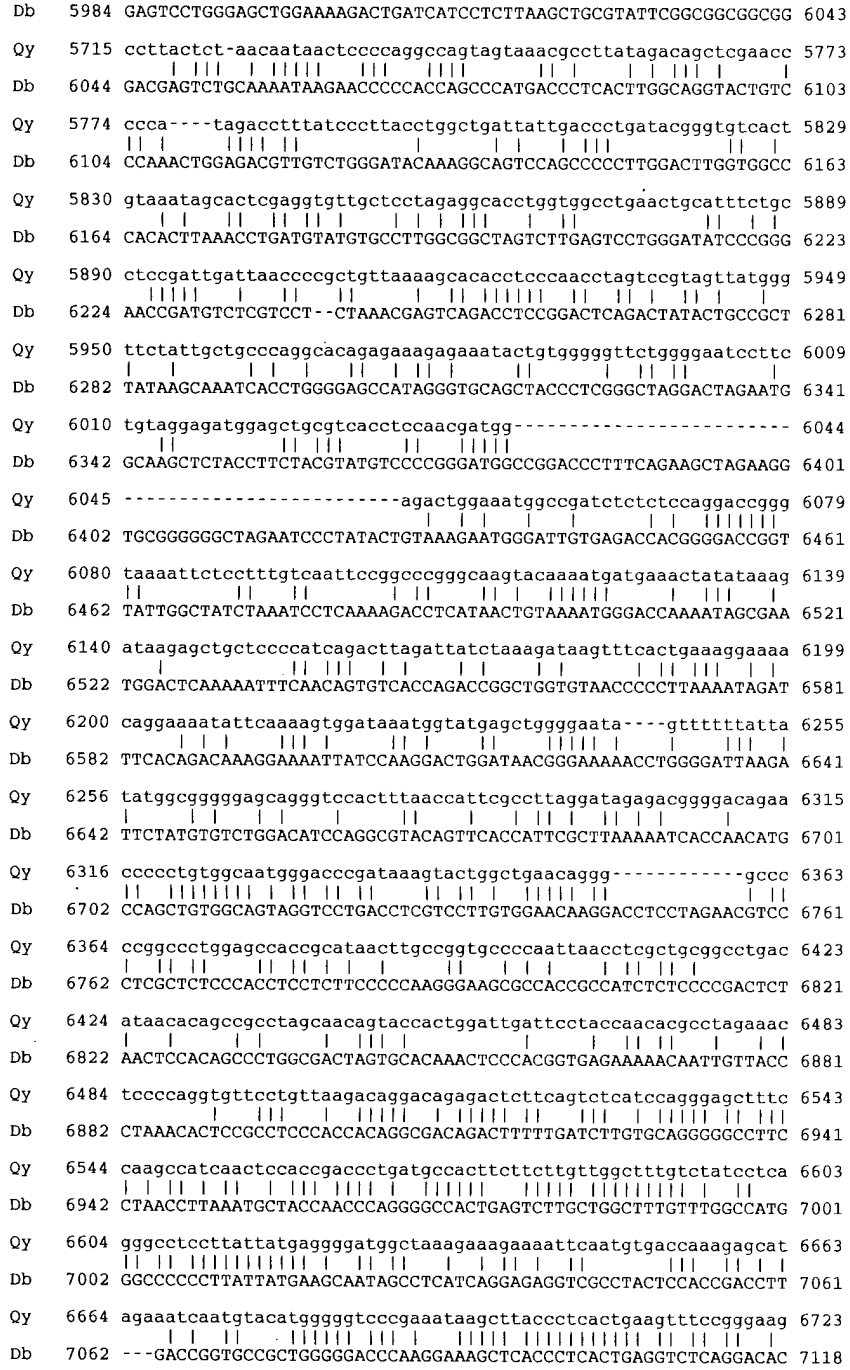
```
Qy   7114 catgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagag 7173
Db   7302 CGTACAGCTATGGATACTGATCTCAGAGCTATAGAACACTCTATAACCAAACTAGAAGAA 7361

Qy   7174 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 7233
Db   7362 TCTTTAACTTCTCTGTCTGAGGTAGTACTGCAAAATAGGAGAGGATTAGATTTGTTATTT 7421

Qy   7234 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 7293
Db   7422 CTTAAAGAAGGAGGACTTTGTGCTGCTCTCAAAGAGGAATGTTGTTTCTATGTCGACCAC 7481

Qy   7294 tcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagg 7353
Db   7482 TCGGGAGTGATCAAAGATTCTATGGCCAAACTTAGAGAACGCCTAGATATACGTCAAAGA 7541

Qy   7354 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 7413
Db   7542 GAGAGAGAGAGCAAACAAGGATGGTTTGAAAGCTGGTTTAATAAGTCCCCCTGGCTCACC 7601

Qy   7414 accctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttggg 7473
Db   7602 ACTCTCCTTTCCACTATAGCGGGACCCTTGATTATACTTCTGCTTTTGCTTACTTTTGGC 7661

Qy   7474 ccttgcttaattaataggtttgttgccttttgttagagaacgagtgagtgcagtccagatc 7533
Db   7662 CCCTGCATTCTTAATAAGTTAGTAGCCTTTATTAGAGAAAGAATAAATGCAGTACAGGTT 7721

Qy   7534 atggtacttaggcaacagtaccaaggccttctg 7566
Db   7722 ATGGTATTAAAACAACAATATCAGGTCTTCCAG 7754
```

```
RESULT   6
US-08-716-351A-1
; Sequence 1, Application US/08716351A
; Patent No. 6033905
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Gibbon Ape Leukemia Virus-Based
;    TITLE OF INVENTION:  Retroviral Vectors
;    NUMBER OF SEQUENCES:  5
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/716,351A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/03784
;      FILING DATE:  06-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian,, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  15280-128-1PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..8535
;      OTHER INFORMATION:  /standard_name= "GaLV SEATO Genome"
```

US-08-716-351A-1

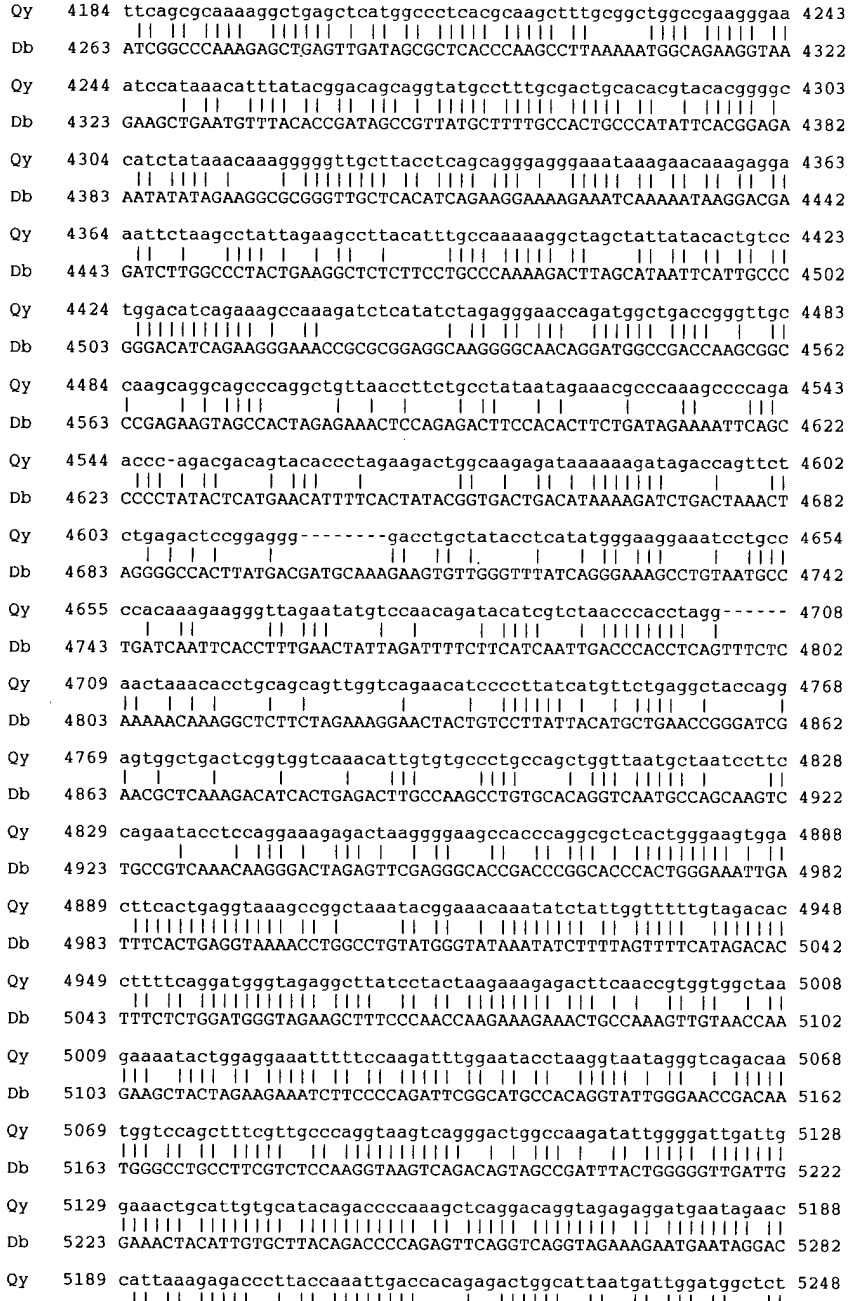
Query Match 27.3%; Score 2244.4; DB 3; Length 8535;
Best Local Similarity 59.1%; Pred. No. 0;
Matches 4366; Conservative 0; Mismatches 2811; Indels 215; Gaps 22;

```
Qy    332 agcagacgtgctaggaggatcacaggctgccaccctgggggacgccccgggaggtgggga 391
Db    756 AACAGACGTGTTCGGGGCTCACCGCCTGGTAATCCAGGGAGACGTCCCAGGATCAGGGGA 815

Qy    392 gagccagggacgcctggtggtctcctactgtcggtcagaggaccgagttctgttgttgaa 451
Db    816 GGACCAGGGACGCCTGGTGGACCCCTCG----------GTAACGGGTCGTTGTGACCCG 864

Qy    452 gcgaaagcttccccctccgcggccgtccgactcttttgcctgcttgtggaagacgcggac 511
Db    865 ATTTCATCGCCCGTCTGGTAAGACGCGCTCTGAATCTGATTCTCTCTCTCGGTCGCCTCG 924

Qy    512 gggtcgcgtgtgtctggatctgttggtttctgtctcgtgtgtctttgtcttgtgcgtcct 571
Db    925 CCGCCGTCTCTGGTTTCTTTTTGTTTCGTTTCTGGAAAGCCTCTGTGTCACAGTCTTTCT 984

Qy    572 tgtctacagttttaatatgggacagacagtgactacccccctttagtttgactctcgacca 631
Db    985 CTCCCAAATCATCAATATGGGACAAGATAATTCTACCCCTATCTCCCTCACTCTAAATCA 1044

Qy    632 ttggactgaagttagatccagggctcataatttgtcagttcaggttaagaagggaccttg 691
Db   1045 CTGGAGAGATGTGAGAACAAGGGCTCACAATCTATCCGTGGAAATCAAAAAGGGAAAATG 1104

Qy    692 gcagactttctgtgcctctgaatggccaacattcgatgttggatggccatcagaggggac 751
Db   1105 GCAGACTTTCTGTTCCTCCGAGTGGCCCACATTCGGCGTGGGGTGGCCACCGGAGGGAAC 1164

Qy    752 ctttaattctgaaattatcctggctgttaaggcaatcatttttcagactggacccggctc 811
Db   1165 TTTTAATCTCTCTGTCATTTTTGCAGTTAAAAAGATTGTCTTTCAG---GAGAACGGGGG 1221

Qy    812 tcatcctgatcaggagccctatatccttacgtggcaagatttggcagaagatcctccgcc 871
Db   1222 ACATCCGGACCAAGTTCCATATATCGTGGTATGGCAGGACCTCGCCCAGAATCCCCCACC 1281

Qy    872 atgggttaaaccatggctaaataaaccaagaaagccaggtccccgaatcctggctcttgg 931
Db   1282 ATGGGTGCCAGCCTCCGCCAAGGTCGCTGTTGTCTCTGATACCCGAAGACCAGTTGCGGG 1341

Qy    932 agagaaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctaccccgaga 991
Db   1342 GAGGCCATCAGCTCCTCCCCGACCCCCCATCTACCCGGCAACAGACGACTTACTCCTCCT 1401

Qy    992 tcgaggagccgccgacttggccggaac--cccaacctgttcccccaccccccttatccagc 1049
Db   1402 CTCTGAACCCACGCCCCCGCCCTATCCGGCGGCACTGCCACCCCCTCTGGCCCCTCAGGC 1461

Qy   1050 acagggtgctgtgaggggacctctgcccc-tcctggagctccggtggtggagggacctgc 1108
Db   1462 GATCGGACCGCCGTCAGGCCAGATGCCCGATAGTAGCGATCCTGAGGGGCCAGCCGCTGG 1521

Qy   1109 tgccgggactcggag---ccggagaggcgccaccccggagcggacagacgagatcgcgat 1165
Db   1522 GACCAGGAGTCGCCGTGCCCGCAGTCCAGCAGACAACTCGGGTCCTGACTCCACTGTGAT 1581

Qy   1166 attaccgctgcgcacctatggccctcccatgccaggggccaattgcagcccctccagta 1225
Db   1582 TTTGCCCCTCCGAGCCATAGGACCCCCGGCCGAGCCCAATGGCCTGGTCCCTCTACAATA 1641

Qy   1226 ttggccctttttcttctgcagatctctataattggaaaactaaccatccccctttctcgga 1285
Db   1642 TTGGCCTTTTTCCTCAGCAGATCTTTATAATTGGAAATCTAATCATCCCTCTTTTTCTGA 1701

Qy   1286 ggatccccaacgcctcacggggttggtggagtcccttatgttctctcaccagcctacttg 1345
```



RESULT 7
US-08-110-300A-8
; Sequence 8, Application US/08110300A
; Patent No. 5643756
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; APPLICANT: Kayman, Samuel

```
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,300A
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-110-300A-8


  Query Match             21.3%;  Score 1747;  DB 1;  Length 8323;
  Best Local Similarity   55.6%;  Pred. No. 0;
  Matches 4058;  Conservative    0;  Mismatches 3005;  Indels  238;  Gaps   27;

Qy    467 tccgcggccgtccgactcttttgcctgcttgtggaa--gacgcggacgggtcgcgtgtgt 524
          |||||  ||||| || | |||     |   | |||||  || || |     |       |||
Db    498 TCCGCCCCCGTCTGAGTTTTTGCTTTCGGTTTGGAACCGAAGCCGCGCCGCGCGTCTTGT 557

Qy    525 ctggatctgttggtttctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagtttt 584
          |||   | |     |||||| | || | | |||| ||  |   | ||    |    |
Db    558 CTGCTGCAGCATCGTTCTGTGTTGTCTCTGTTTGACTGTT--TTTCTGTATTTGTCTGAA 615

Qy    585 aatatgggacagacagtgactacccccctagtttgactctcgaccattggactgaagtt 644
          || ||||| ||| | || || |||||| | |||||||||| | |||||  ||||  || ||
Db    616 AACATGGGCCAGGCTGTTACCACCCCCTTAAGTTTGACTTTAGACCACTGGAAGGATGTC 675

Qy    645 agatccagggctcataatttgtcagttcaggttaagaagggaccttggcagactttctgt 704
            |   | || || ||  |||| ||  ||||||  ||  | |  |||   || |||||
Db    676 GAACGGACAGCCCACAACCTGTCGGTAGAGGTTAGAAAAAGGCGCTGGGTTACATTCTGC 735

Qy    705 gcctctgaatggccaacattcgatgttggatggccatcagaggggacctttaattctgaa 764
            | |||||||||||| ||| | || |||||||||  ||| || || |||||  | ||
Db    736 TCTGCAGAATGGCCAACCTTCAACGTCGGATGGCCACGAGACGGCACTTTTAACCCAGAC 795

Qy    765 attatcctggctgttaaggcaatcatttttcagactggacccggctctcatcctgatcag 824
          |||||       ||||||   |   | ||     |||| ||       ||||| ||||||
Db    796 ATTATTACACAGGTTAAGATCAAGGTCTTCTCACCTGGCCCACATGGACATCCGGATCAG 855

Qy    825 gagccctatatccttacgtggcaagatttggcagaagatcctccgccatgggttaaacca 884
          |  ||||| ||| | || ||| ||| | | |||| ||| || || || ||||| | |||
Db    856 GTCCCCTACATCGTGACCTGGGAAGCTATAGCAGTAGACCCCCCTCCCTGGGTCAGACCC 915

Qy    885 tggctaaa-------------------taaaccaagaaagccaggtccccgaatcctggc 925
          |   |  |                    ||     ||||   || |     ||||
Db    916 TTCGTGCACCCTAAACCTCCCCTCTCTCTTCCCCCTTCAGCCCCCTCTCTCCCACCTGAA 975

Qy    926 tcttggagagaaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctacc 985
          |           |  |   | |   |         ||  || |||         |  | |
Db    976 CCCCCACTCTCGACCCCGCCCCAGTCCTCCCTCTATCCGGCTCTCACTTCTCCTTTAAAC 1035

Qy    986 ccgagatcgaggagccgccgacttggccggaaccccaacctgttcccccacccccttatc 1045
          | | | | |||   |     |||       | |       |   ||    | |  || |
Db   1036 ACCAAACCTAGGCCTCAAGTCCTTCCTGATAGCGGAGGACCACTCATTGATCTACTCACG 1095

Qy   1046 cagcacagggtgctgtgaggggacctctgcccctcctggagctccggtggtggagggacc 1105
          || ||    ||  |  ||||||   | ||  |||       |||    || |  | |
Db   1096 GAGGACCCTCCGCCTTACCGGGACCCAGGGCCACCCTCTCCTGACGGGAACGGCGATAGC 1155

Qy   1106 tgctgccgggactc-------ggagccggagaggcgccacc----------------ccg 1142
          |      || | |       ||||||   ||  |  | ||                 | |
Db   1156 GGAGAAGTGGCCCCTACAGAAGGAGCCCCTGACCCTTCCCCAATGGTATCCCGCCTGCGG 1215

Qy   1143 gagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgccaggg 1202
          |   | | |||      || |       |      | |  | |     ||| | |    |
Db   1216 GGAAGAAAAGAACCCCCCGTGGCGGATTCTACTACCTCTCAGGCGTTCCCCCTTCGCCTG 1275

Qy   1203 ggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattggaaa 1262
          ||          |     || || ||||| ||||| || | || |||||||| ||||||
Db   1276 GGAGGGAATGGACAGTATCAATACTGGCCATTTTCCTCCTCTGACCTCTATAACTGGAAA 1335

Qy   1263 actaaccatcccccttttctcggaggatccccaacgcctcacggggttggtggagtccctt 1322
          | |||| | ||| ||||||| ||||| ||        | || |  ||| | |||||| ||
Db   1336 AATAACAACCCCTCTTTCTCCGAGGACCCAGCTAAATTGACAGCTTTGATCGAGTCCGTT 1395

Qy   1323 atgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttcaca 1382
          |  |  |||| ||||| |||||||||||| || ||||||||  |   ||| ||  | ||
Db   1396 CTCCTTACTCATCAGCCCACTTGGGATGACTGCCAACAGCTATTAGGGACCCTGCTGACG 1455

Qy   1383 accgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgacggg 1442
            || ||   | || || | ||  |||||||  ||||    ||||  ||||  |||||
Db   1456 GGAGAAGAAAAACAGCGAGTGCTCCTAGAGGCCCGAAAGGCGGTTCGAGGGGAGGACGGA 1515

Qy   1443 cgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggttgg 1502
          || || || ||| |||  ||||| ||| |    |  |||||||||   || ||||  |||
Db   1516 CGCCCAACTCAGCTGCCCAATGACATTAATGATGCTTTTCCCTTGGAACGTCCCGACTGG 1575

Qy   1503 gactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcg 1562
          ||||||||||       |||||||  |   | |      |||||||||   ||  | |||
Db   1576 GACTACAACACCCAACGAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCG 1635

Qy   1563 ggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcag 1622
          |||||||    |||    ||| | ||||| |||||||| |||||||  || | |  |||
Db   1636 GGTCTCCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGGATAACCCAG 1695

Qy   1623 ggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcggttc 1682
          ||||| || ||  |||||||  |  ||| | ||||| |||| ||| ||||   |  | | |
Db   1696 GGACCTAATGAGTCTCCCTCAGCCTTTTTAGAGAGACTCAAGGAGGCCTATCGCAGATAC 1755

Qy   1683 acccctttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcattggg 1742
          || |||| ||| |||          |  ||  || ||   ||||||  || | |||||  ||
Db   1756 ACTCCTTATGACCCTGAGGACCCAGGGCAAGAAACCAATGTGGCCATGTCATTCATCTGG 1815

Qy   1743 cagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggctgag 1802
          ||||| || | |||||||  ||   ||  | || |  | ||||  ||  ||
Db   1816 CAGTCCGCCCCGGATATCGGGCGAAAGTTAGAGCGGTTAGAAGATTTGAAGAGTAAGACC 1875

Qy   1803 ttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagaggag 1862
          ||| | ||  |||||||  || || || ||| | | | | | | | || ||    || ||
Db   1876 TTAGGAGACTTAGTGAGGGAAGCTGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAA 1935

Qy   1863 aaggaacagagaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacggcaa 1922
          |  ||  |  | |     |  || | ||| |||| ||| ||| | |||     |  | |
Db   1936 AGAGAGGAACGTATTAGGAGAGAAACAGAGGAAAAGGAAGAACGCCGTAGGGCAGAGGAT 1995

Qy   1923 gagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggagaga 1982
          | | |||    ||   |||       | |      | ||  | ||| |           |
Db   1996 GTGCAGAGAGAGAAGGAGAGGGACCGCAGAAGACATAGAGAAATGAGTA----------A 2045

Qy   1983 gagagagattttaggaaaattaggtcaggccctagacagtcaggga-acctgggcaatag 2041
```

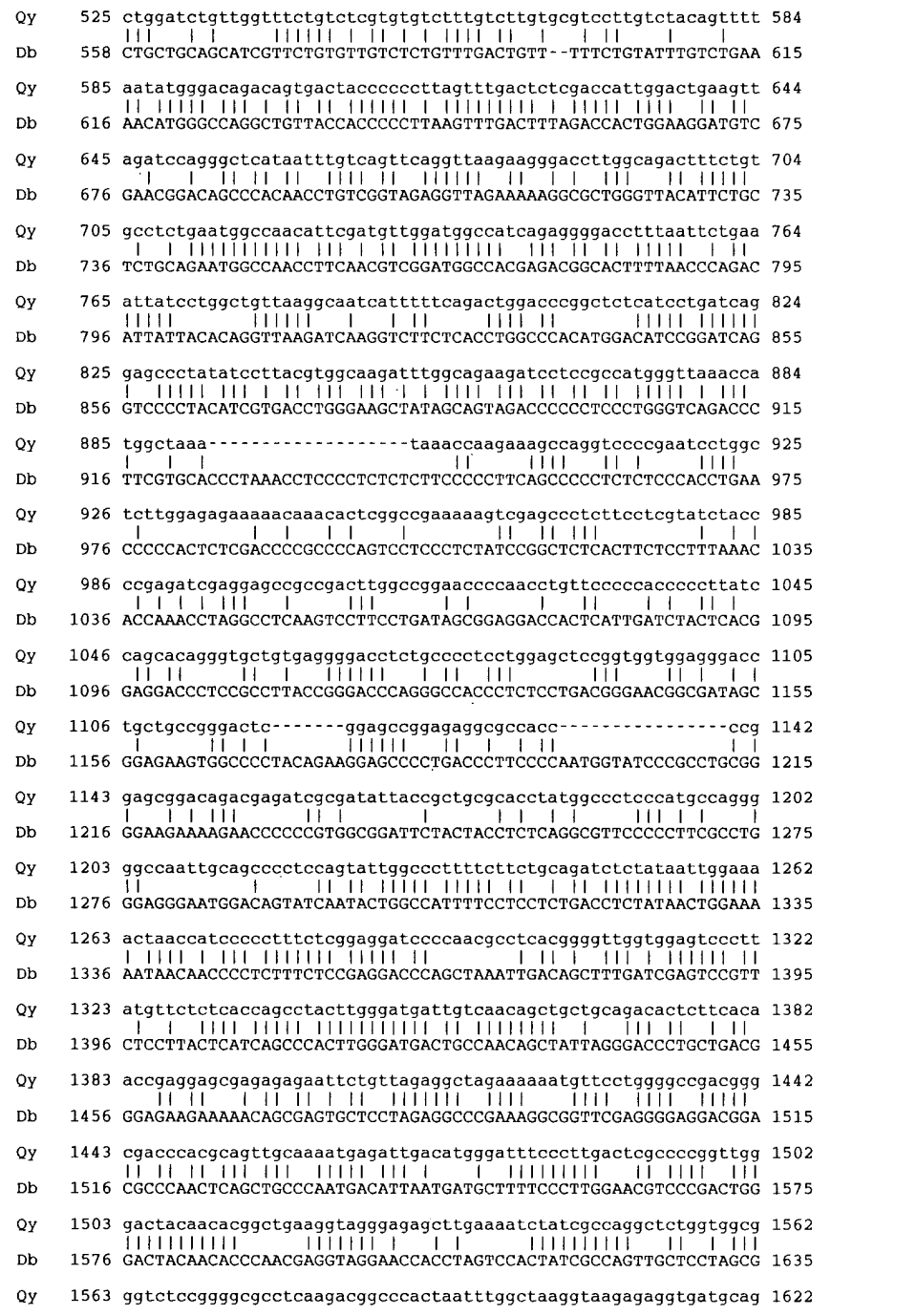

Qy 4184 ttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaagggaa 4243
Db 4263 ATCGGCCCAAAGAGCTGAGTTGATAGCGCTCACCCAAGCCTTAAAAATGGCAGAAGGTAA 4322
Qy 4244 atccataaacatttatacggacagcaggtatgcctttgcgactgcacacgtacacggggc 4303
Db 4323 GAAGCTGAATGTTTACACCGATAGCCGTTATGCTTTTGCCACTGCCCATATTCACGGAGA 4382
Qy 4304 catctataaacaaaggggttgcttacctcagcagggagggaaataaagaacaaagagga 4363
Db 4383 AATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAAATCAAAAATAAGGACGA 4442
Qy 4364 aattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgtcc 4423
Db 4443 GATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGACTTAGCATAATTCATTGCCC 4502
Qy 4424 tggacatcagaaagccaaagatctcatatctagagggaaccagatggctgaccgggttgc 4483
Db 4503 GGGACATCAGAAGGGAAACCGCGCGGAGGCAAGGGGCAACAGGATGGCCGACCAAGCGGC 4562
Qy 4484 caagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaaagccccaga 4543
Db 4563 CCGAGAAGTAGCCACTAGAGAAACTCCAGAGACTTCCACACTTCTGATAGAAAATTCAGC 4622
Qy 4544 accc-agacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttct 4602
Db 4623 CCCCTATACTCATGAACATTTTCACTATACGGTGACTGACATAAAAGATCTGACTAAACT 4682
Qy 4603 ctgagactccggaggg--------gacctgctatacctcatatgggaaggaaatcctgcc 4654
Db 4683 AGGGGCCACTTATGACGATGCAAAGAAGTGTTGGGTTTATCAGGGAAAGCCTGTAATGCC 4742
Qy 4655 ccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctagg------ 4708
Db 4743 TGATCAATTCACCTTTGAACTATTAGATTTTCTTCATCAATTGACCCACCTCAGTTTCTC 4802
Qy 4709 aactaaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccagg 4768
Db 4803 AAAAACAAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTACATGCTGAACCGGGATCG 4862
Qy 4769 agtggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttc 4828
Db 4863 AACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAGGTCAATGCCAGCAAGTC 4922
Qy 4829 cagaatacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtgga 4888
Db 4923 TGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCGACCCGGCACCCACTGGGAAATTGA 4982
Qy 4889 cttcactgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacac 4948
Db 4983 TTTCACTGAGGTAAAACCTGGCCTGTATGGGTATAAATATCTTTTAGTTTTCATAGACAC 5042
Qy 4949 cttttcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaa 5008
Db 5043 TTTCTCTGGATGGGTAGAAGCTTTCCCAACCAAGAAAGAAACTGCCAAAGTTGTAACCAA 5102
Qy 5009 gaaaatactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaa 5068
Db 5103 GAAGCTACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAGGTATTGGGAACCGACAA 5162
Qy 5069 tggtccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattg 5128
Db 5163 TGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGATTTACTGGGGGTTGATTG 5222
Qy 5129 gaaactgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaac 5188
Db 5223 GAAACTACATTGTGCTTACAGACCCCAGAGTTCAGGTCAGGTAGAAAGAATGAATAGGAC 5282
Qy 5189 cattaaagagacccttaccaaattgaccacagagactggcattaatgattggatggctct 5248
Db 5283 AATCAAGGAGACTTTAACTAAATTGACGCTTGCAACTGGCTCTAGGGACTGGGTGCTCCT 5342
Qy 5249 cctgccctttgtgcttttagggtgaggaacacccctggacagtttgggctgaccccta 5308
Db 5343 GCTTCCCCTAGCCCTGTATCGAGCCCGCAACACGCCGGGCCCCCATGGTCTCACCCCATA 5402
Qy 5309 tgaattgctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctga 5368
Db 5403 TGAAATCTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGATCCTGACATGGCAAA 5462
Qy 5369 tgtgctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcg 5428
Db 5463 GGTTACTCATAACCCCTCTCTCCAAGCCCATTTACAGGCACTCTACCTGGTCCAGCACGA 5522
Qy 5429 agcgtggaagcagctccgggaggcctactcagga------ggagacttgcaagttccaca 5482
Db 5523 AGTCTGGAGACCGTTGGCGGCAGCTTACCAAGAACAACTGGACCGGCCGGTAGTGCCTCA 5582
Qy 5483 tcgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagac 5542
Db 5583 CCCTTTCCGAGTCGGTGACACAGTGTGGGTCCGCAGACACCAAACTAAAAATCTAGAACC 5642
Qy 5543 tcggtggaagggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaagg 5602
Db 5643 CCGCTGGAAAGGACCTTATACCGTCCTACTGACTACCCCCACCGCTCTCAAAGTGGACGG 5702
Qy 5603 aatccccacctggatccatgcatcccacgttaagccggc----------gccacc 5647
Db 5703 CATTGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGACACCAGGATTGAGCCACC 5762
Qy 5648 tcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcgcctccatcg 5707
Db 5763 ATCGGAATCGACATGGCGTGTTCAACGCTCTCAAAATCCCCTAAAGATAAGATTGACCCG 5822
Qy 5708 cctggttccttactctaacaataactccccaggccagtagtaaacgccttatagacagct 5767
Db 5823 CGGGACCTCCTAATCCCCTTAATTCTCTTCCTGTCTCTCAAAGGGCCAGATCCGCAGCA 5882
Qy 5768 cgaaccccca----tagacctttatcccttacctggctgattattgaccctgatacgggt 5823
Db 5883 CCCGGCTCCAGCCCTCACCAGGTCTACAACATTACCTGGGAAGTGACCAATGGGGATCGG 5942
Qy 5824 gtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac----- 5878
Db 5943 GAGACAGTATGGGCAATATCAGGCAACCACCCTCTGTGGACTTGGTGGCCAGTCCTCACC 6002
Qy 5879 ---------tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctccca 5929
Db 6003 CCAGATTTGTGTATGTTAGCTCTCAGTGGGCCGCCCCACTGGGGGCTAGAGTATCAGGCC 6062
Qy 5930 acctagtccgtagttatgggttc----tattgctgcccaggcacagagaaagagaaatac 5985
Db 6063 CCCTATTCCTCGCCCCCGGGGCCCCCTTGTTGCTCAGGGAGCAGCGGGAACGTTGCAGGC 6122
Qy 5986 tgtggggg----------ttctggggaatccttctgtagga 6016
Db 6123 TGTGCCAGAGACTGCAACGAGCCCTTGACCTCCCTCACCCCTCGGTGCAACACTGCCTGG 6182
Qy 6017 gatggagctgcgtcacctccaacgatggagactggaaatggccgatctctctccaggacc 6076
Db 6183 AACAGACTTAAGCTGGACCAGGTAACTCATAAATCAAGTGAGGGATTTTATGTCTGCCCC 6242
Qy 6077 gggtaaaattctcctttgtcaattccggcccgggcaagtacaaaatgatgaaact----- 6131
Db 6243 GGGTCACATCGCCCCCGGGAAGCCAAGTCCTGTGGGGGTCCAGACTCCTTCTACTGTGCC 6302
Qy 6132 -------------------------atataaagataagagctgctccccatcagactta 6165
Db 6303 TCTTGGGGCTGCGAGACAACCGGTAGAGTATACTGGAAGCCCTCCTCTTCTTGGGACTAC 6362
Qy 6166 gattatctaaagataagtttcactgaaaggaaaacaggaaaatattcaaaagtggataaa 6225
Db 6363 ATCACAGTAGACAACAATCTCACCTCTAACCAGGCTGTTCAGGTATGCAAAGACAATAAG 6422
Qy 6226 tggtatgagctggggaatagttttttattatatggcgggggagcagggtccactttaacc 6285

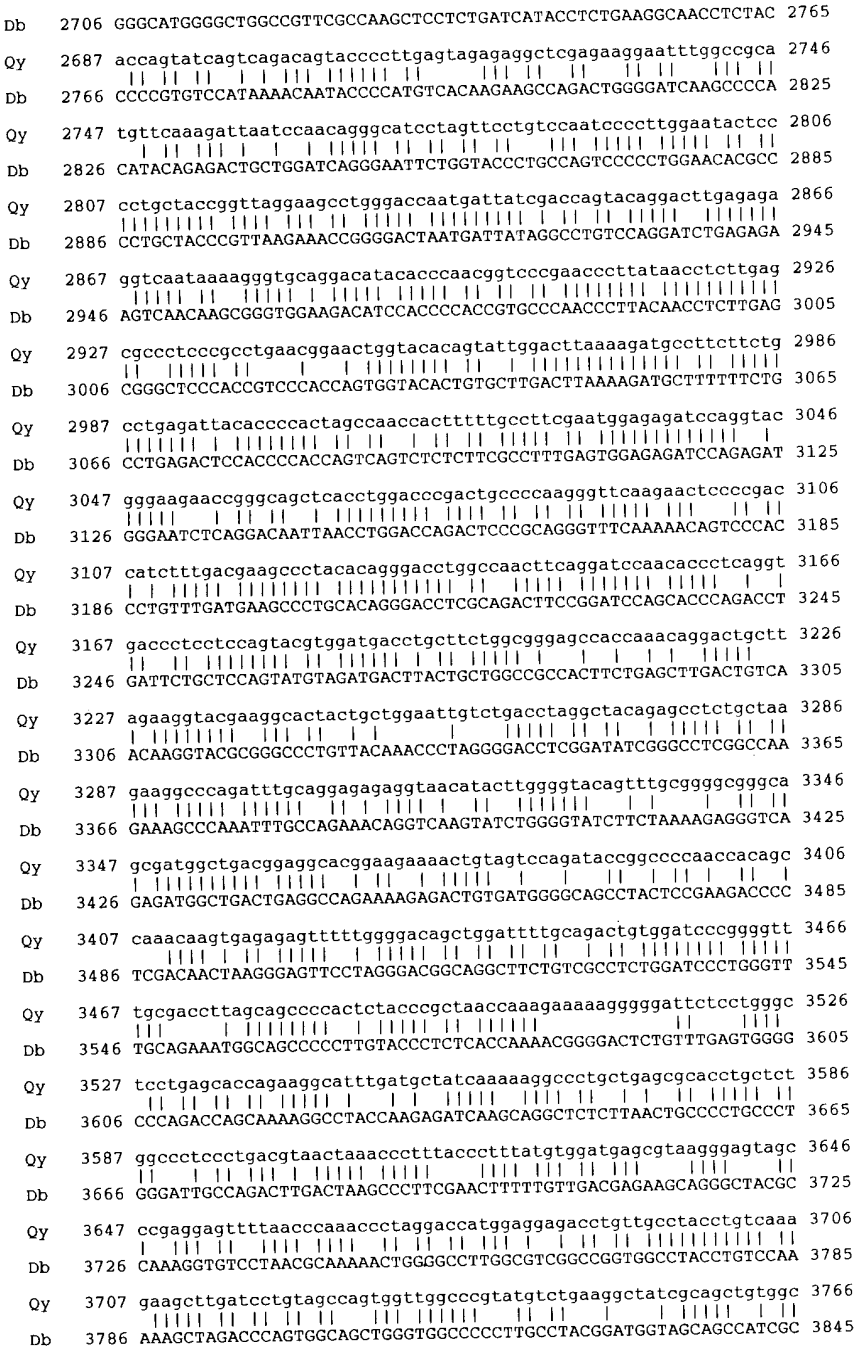
```
Db   6423 TGGTGCAATCCCTTGGCTATCCGGTTTACAAACGCCGGGAAACAGGTCACCTCATGGACA 6482

Qy   6286 a---------------ttcgccttaggatagagacggggacagaaccccctgtggcaatg 6330
Db   6483 ACTGGACACTATTGGGGTCTACGTCTTTATGTCTCTGGACAGGACCCAGGGCTTACTTTC 6542

Qy   6331 ggacccgataaagtactggctgaacaggggccccggccctggagccaccgcataacttg 6390
Db   6543 GGGATCCGACTCAGTTATCAAAATCTAGGACCTCGGATCCCAATAGGACCAAACCCCGTC 6602

Qy   6391 ccggtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtacc 6450
Db   6603 CTGGCAGACCAACTTTCGTTCCCGCTACCTAATCCCCTACCCAAACCTGCCAAGTCTCCC 6662

Qy   6451 actggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaag------ 6504
Db   6663 CCCGCCTCTAGTTCGACTCCCACATTGATTTCCCCGTCCCCCACTCCCACTCAGCCCCCG 6722

Qy   6505 ---------acaggacagagactcttcagtctcatccagggagctttccaagccatcaac 6555
Db   6723 CCAGCAGGAACGGGAGACAGATTACTAAATCTAGTACAGGGAGCTTACCAGGCACTCAAC 6782

Qy   6556 tccaccgaccctgatgccacttcttcttgttggctttgtctatcctcagggcctccttat 6615
Db   6783 CTTACCAACCCTGATAAAACTCAAGAGTGCTGGTTATGCCTAGTGTCTGGACCCCCCTAT 6842

Qy   6616 tatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgt 6675
Db   6843 TACGAGGGGGTTGCCGTCCTAGGTACTTATTCCAACCATACCTCTGCCCCAGCTAACTGC 6902

Qy   6676 acatgggggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgcata 6735
Db   6903 TCCGTGGCCTCCCAACACAAGCTGACCCTGTCCGAAGTGACTGGACGGGGACTCTGCATA 6962

Qy   6736 ggaaaagctcccccatcccaccaacacctttgctatagtactgtggtttatgagcaggcc 6795
Db   6963 GGAACAGTCCCAAAAACTCACCAGGCCCTGTGCAACACTACCCT------TAAGGCAGGC 7016

Qy   6796 tcagaaaatcagtatttagtacctggttataacaggtggtgggcatgcaatactgggtta 6855
Db   7017 AAAGGGTCTTACTATCTAGTTGCCCCCACAGGAACTATGTGGGCATGTAACACTGGACTC 7076

Qy   6856 acccccctgtgtttccacctcagtcttcaaccaatccaaagatttctgtgtcatggtccaa 6915
Db   7077 ACTCCATGCCTATCTGCCACCGTGCTTAATCGCACCACTGACTATTGCGTTCTCGTGGAA 7136

Qy   6916 atcgtcccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcgg 6975
Db   7137 TTATGGCCCAGGGTCACCTACCATCCTCCCAGTTACGTCTATAGCCAGTTTGAAAAATCC 7196

Qy   6976 tataaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggattaggg 7035
Db   7197 CATA---GACATAAAAGAGAACCAGTGTCCTTAACCTTGGCCTTATTATTAGGTGGGCTA 7253

Qy   7036 acggccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagag 7095
Db   7254 ACTATGGGTGGCATCGCCGCGGGAGTAGGGACAGGAACTACCG---CCCTGGTCGCCACC 7310

Qy   7096 aaaggacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtct 7155
Db   7311 CAGCAGTTTCAGCAGCTCCATGCTGCCGTACAAGATGATCTCAAAGAAGTCGAAAAGTCA 7370

Qy   7156 gttagcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggagg 7215
Db   7371 ATTACTAACCTAGAAAAGTCTCTTACTTCGTTGTCTGAGGTTGTACTGCAGAATCGACGA 7430

Qy   7216 ggattagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgt 7275
Db   7431 GGCCTAGACCTGTTGTTCCTAAAAGAGGGAGGACTGTGTGCTGCCCTAAAAGAAGAATGT 7490

Qy   7276 tgcttctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaag 7335

Db   7491 TGTTTCTATGCTGACCATACAGGCCTAGTAAGAGATAGTATGGCCAAATTAAGAGAGAGA 7550

Qy   7336 ttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaac 7395
Db   7551 CTCTCTCAGAGACAAAAACTATTTGAGTCGAGCCAAGGATGGTTCGAAGGATGGTTTAAC 7610

Qy   7396 aggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctc 7455
Db   7611 AGATCCCCCTGGTTTACCACGTTGATATCCACCATCATGGGGCCTCTCATTATACTCCTA 7670

Qy   7456 ctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacga 7515
Db   7671 CTAATTCTGCTTTTTGGACCCTGCATTCTTAATCGATTAGTTCAATTTGTTAAAGACAGG 7730

Qy   7516 gtgagtgcagtccagatcatggtacttaggcaacagtacca 7556
Db   7731 ATCTCAGTAGTCCAGGCTTTAGTCCTGACTCAACAATACCA 7771
```

RESULT 8
US-08-886-642-8
; Sequence 8, Application US/08886642
; Patent No. 5952474
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 45 Rockefeller Plaza, Suite 2800
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/886,642
; FILING DATE: 01-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/110,300
; FILING DATE: 20-AUG-1993
; APPLICATION NUMBER: 07/938,100
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hone, William J.
; REGISTRATION NUMBER: 26,739
; REFERENCE/DOCKET NUMBER: 07763/010002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212/765-5070
; TELEFAX: 212/258-2291
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-886-642-8

Query Match 21.3%; Score 1747; DB 2; Length 8323;
Best Local Similarity 55.6%; Pred. No. 0;
Matches 4058; Conservative 0; Mismatches 3005; Indels 238; Gaps 27;

```
Qy    467 tccgcggccgtccgactcttttgcctgcttgtggaa--gacgcggacgggtcgcgtgtgt 524
Db    498 TCCGCCCCCGTCTGAGTTTTTGCTTTCGGTTTGGAACCGAAGCCGCGCCGCGCGTCTTGT 557
```

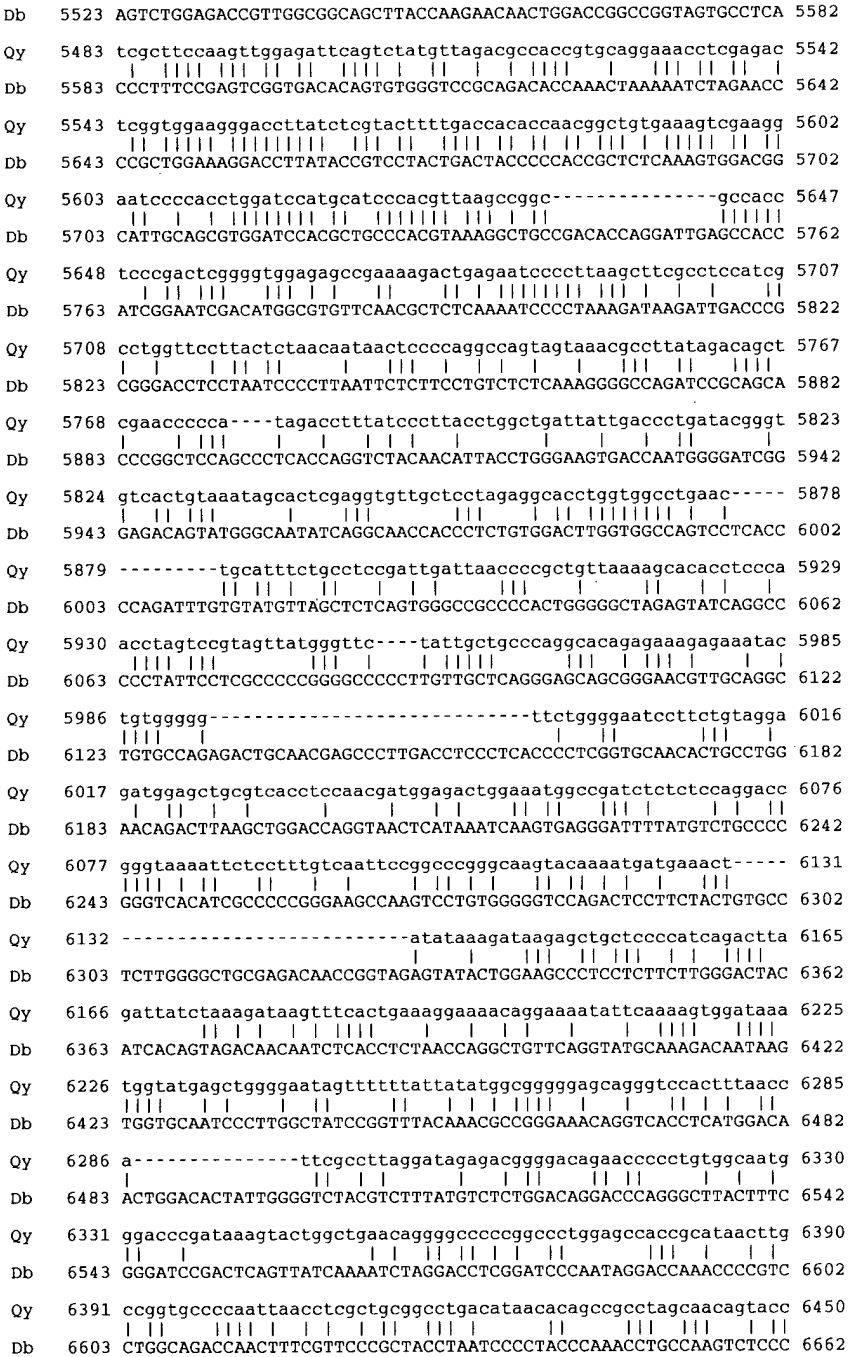

Qy 525 ctggatctgttggtttctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagtttt 584
Db 558 CTGCTGCAGCATCGTTCTGTGTTGTCTCTGTTTGACTGTT--TTTCTGTATTTGTCTGAA 615
Qy 585 aatatgggacagacagtgactacccccctagtttgactctcgaccattggactgaagtt 644
Db 616 AACATGGGCCAGGCTGTTACCACCCCCTTAAGTTTGACTTTAGACCACTGGAAGGATGTC 675
Qy 645 agatccagggctcataatttgtcagttcaggttaagaagggaccttggcagactttctgt 704
Db 676 GAACGGACAGCCCACAACCTGTCGGTAGAGGTTAGAAAAAGGCGCTGGGTTACATTCTGC 735
Qy 705 gcctctgaatggccaacattcgatgttggatggccatcagaggggacctttaattctgaa 764
Db 736 TCTGCAGAATGGCCAACCTTCAACGTCGGATGGCCACGAGACGGCACTTTTAACCCAGAC 795
Qy 765 attatcctggctgttaaggcaatcatttttcagactggacccggctctcatcctgatcag 824
Db 796 ATTATTACACAGGTTAAGATCAAGGTCTTCTCACCTGGCCCACATGGACATCCGGATCAG 855
Qy 825 gagccctatatccttacgtggcaagatttggcagaagatcctccgccatgggttaaacca 884
Db 856 GTCCCCTACATCGTGACCTGGGAAGCTATAGCAGTAGACCCCCCTCCCTGGGTCAGACCC 915
Qy 885 tggctaaa----------taaaccaagaaagccaggtccccgaatcctggc 925
Db 916 TTCGTGCACCCTAAACCTCCCCTCTCTCTTCCCCCTTCAGCCCCCTCTCTCCCACCTGAA 975
Qy 926 tcttggagagaaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctacc 985
Db 976 CCCCCACTCTCGACCCCGCCCCAGTCCTCCCTCTATCCGGCTCTCACTTCTCCTTTAAAC 1035
Qy 986 ccgagatcgaggagccgccgacttggccggaaccccaacctgttcccccaccccccttatc 1045
Db 1036 ACCAAACCTAGGCCTCAAGTCCTTCCTGATAGCGGAGGACCACTCATTGATCTACTCACG 1095
Qy 1046 cagcacagggtgctgtgaggggacctctgcccctcctggagctccggtggtggagggacc 1105
Db 1096 GAGGACCCTCCGCCTTACCGGGACCCAGGGCCACCCTCTCCTGACGGGAACGGCGATAGC 1155
Qy 1106 tgctgccgggactc-------ggagccggagaggcgccacc----------ccg 1142
Db 1156 GGAGAAGTGGCCCCTACAGAAGGAGCCCCTGACCCTTCCCCAATGGTATCCCGCCTGCGG 1215
Qy 1143 gagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatgccaggg 1202
Db 1216 GGAAGAAAAGAACCCCCCGTGGCGGATTCTACTACCTCTCAGGCGTTCCCCCTTCGCCTG 1275
Qy 1203 ggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattggaaa 1262
Db 1276 GGAGGGAATGGACAGTATCAATACTGGCCATTTTCCTCCTCTGACCTCTATAACTGGAAA 1335
Qy 1263 actaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtccctt 1322
Db 1336 AATAACAACCCCTCTTTCTCCGAGGACCCAGCTAAATTGACAGCTTTGATCGAGTCCGTT 1395
Qy 1323 atgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttcaca 1382
Db 1396 CTCCTTACTCATCAGCCCACTTGGGATGACTGCCAACAGCTATTAGGGACCCTGCTGACG 1455
Qy 1383 accgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgacggg 1442
Db 1456 GGAGAAGAAAAACAGCGAGTGCTCCTAGAGGCCCGAAAGGCGGTTCGAGGGGAGGACGGA 1515
Qy 1443 cgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggttgg 1502
Db 1516 CGCCCAACTCAGCTGCCCAATGACATTAATGATGCTTTTCCCTTGGAACGTCCCGACTGG 1575
Qy 1503 gactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcg 1562
Db 1576 GACTACAACACCCAACGAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCG 1635
Qy 1563 ggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcag 1622
Db 1636 GGTCTCCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGGATAACCCAG 1695
Qy 1623 ggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcggttc 1682
Db 1696 GGACCTAATGAGTCTCCCTCAGCCTTTTAGAGAGACTCAAGGAGGCCTATCGCAGATAC 1755
Qy 1683 acccctttgatcctacctcagaggcccagaaagcctcagtggccctggccttcattggg 1742
Db 1756 ACTCCTTATGACCCTGAGGACCCAGGGCAAGAAACCAATGTGGCCATGTCATTCATCTGG 1815
Qy 1743 cagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggctgag 1802
Db 1816 CAGTCCGCCCCGGATATCGGGCGAAAGTTAGAGCGGTTAGAAGATTTGAAGAGTAAGACC 1875
Qy 1803 ttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagaggag 1862
Db 1876 TTAGGAGACTTAGTGAGGGAAGCTGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAA 1935
Qy 1863 aaggaacagagaaaagaaaaggagagagaagaaaggggaggaaagacgtgatagacggcaa 1922
Db 1936 AGAGAGGAACGTATTAGGAGAGAAACAGAGGAAAAGGAAGAACGCCGTAGGGCAGAGGAT 1995
Qy 1923 gagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggagaga 1982
Db 1996 GTGCAGAGAGAGAAGGAGAGGGACCGCAGAAGACATAGAGAAATGAGTA----------A 2045
Qy 1983 gagagagattttaggaaaattaggtcaggccctagacagtcaggga-acctgggcaatag 2041
Db 2046 GTTGCTGGCTACTGTCGTTAGCGGGCAGAGACAGGATAGACAGGGAGGAGAGCGAAGGAG 2105
Qy 2042 gaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaaggaa 2101
Db 2106 GCCCCAACTCGACCACGACCAGTGTGCCTACTGCAAAGAAAAGGGACATTGGGCTAGAGA 2165
Qy 2102 ctgccccaagaagggaaacaaaggaccgaaggtccta---------gctctagaagaaga 2152
Db 2166 TTGCCCCAAGAAGCCAAGAGGACCCCGGGGACCACGACCCCAGGCCTCCCTCCTGACCTT 2225
Qy 2153 taaagattaggggagacggggttcggaccccctccccgagcccagggtaactttgaaggt 2212
Db 2226 AGACGATTAGGGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAGAGT 2285
Qy 2213 ggaggggcaaccagttgagttcctggttgataccggagcggagcattcagtgctgctaca 2272
Db 2286 CGGGGGGCAACCCGTCACCTTCCTAGTGGATACTGGGGCCCAACACTCCGTGCTGACCCA 2345
Qy 2273 accattaggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggca 2332
Db 2346 AAATCCTGGACCCCTAAGTGACAAGTCTGCCTGGGTCCAAGGGGCTACTGGAGGGAAGCG 2405
Qy 2333 gtatccatggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtt 2392
Db 2406 GTATCGCTGGACCACGGATCGCCGAGTGCACCTAGCCACCGGTAAGGTCACCCATTCTTT 2465
Qy 2393 tctggtcatccctgagtgcccagtacccctttctaggtagagacttactgaccaagatggg 2452
Db 2466 CCTCCATGTACCAGATTGCCCCTATCCTCTGCTAGGAAGAGATTTGCTGACTAAACTAAA 2525
Qy 2453 agctcaaatttcttttga---acaaggaagaccagaagtgtctgtgaataacaaacccat 2509
Db 2526 AGCCCAAATTCACTTTGAGGGATCAGGAGCTCAGGTTGTGGGACCAATGGGACAGCCCCT 2585
Qy 2510 cactgtgttgaccctccaattagatgatgaatatcgactatattctccccaagtaaagcc 2569
Db 2586 GCAAGTGCTGACCCTAAACATAGAAGATGAGTATCGGCTACATGAGACCTCAAAAGGGCC 2645
Qy 2570 tgatcaagatatacagtcc---tggttggagcagtttccccaagcctgggcagaaaccgc 2626
Db 2646 AGATGTGCCTCTAGGGTCCACATGGCTCTCTGATTTTCCCCAGGCCTGGGCAGAAACCGG 2705
Qy 2627 agggatgggtttggcaaagcaagttcccccacaggttattcaactgaaggccagtgctac 2686


Db 2706 GGGCATGGGGCTGGCCGTTCGCCAAGCTCCTCTGATCATACCTCTGAAGGCAACCTCTAC 2765
Qy 2687 accagtatcagtcagacagtacccctgagtagagaggctcgagaaggaatttggccgca 2746
Db 2766 CCCCGTGTCCATAAAACAATACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCA 2825
Qy 2747 tgttcaaagattaatccaacagggcatcctagttcctgtccaatccccttggaatactcc 2806
Db 2826 CATACAGAGACTGCTGGATCAGGGAATTCTGGTACCCTGCCAGTCCCCCTGGAACACGCC 2885
Qy 2807 cctgctaccggttaggaagcctgggaccaatgattatcgaccagtacaggacttgagaga 2866
Db 2886 CCTGCTACCCGTTAAGAAACCGGGGACTAATGATTATAGGCCTGTCCAGGATCTGAGAGA 2945
Qy 2867 ggtcaataaaagggtgcaggacatacacccaacggtcccgaacccttataacctcttgag 2926
Db 2946 AGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCCAACCCTTACAACCTCTTGAG 3005
Qy 2927 cgccctcccgcctgaacggaactggtacacagtattggacttaaaagatgccttcttctg 2986
Db 3006 CGGGCTCCCACCGTCCCACCAGTGGTACACTGTGCTTGACTTAAAAGATGCTTTTTTCTG 3065
Qy 2987 cctgagattacaccccactagccaaccacttttttgccttcgaatggagagatccaggtac 3046
Db 3066 CCTGAGACTCCACCCCACCAGTCAGTCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGAT 3125
Qy 3047 gggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactccccgac 3106
Db 3126 GGGAATCTCAGGACAATTAACCTGGACCAGACTCCCGCAGGGTTTCAAAAACAGTCCCAC 3185
Qy 3107 catctttgacgaagccctacacagggacctggccaacttcaggatccaacaccctcaggt 3166
Db 3186 CCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGGATCCAGCACCCAGACCT 3245
Qy 3167 gaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggactgctt 3226
Db 3246 GATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCCACTTCTGAGCTTGACTGTCA 3305
Qy 3227 agaaggtacgaaggcactactgctggaattgtctgacctaggctacagagcctctgctaa 3286
Db 3306 ACAAGGTACGCGGGCCCTGTTACAAACCCTAGGGGACCTCGGATATCGGGCCTCGGCCAA 3365
Qy 3287 gaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggcgggca 3346
Db 3366 GAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCA 3425
Qy 3347 gcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaaccacagc 3406
Db 3426 GAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 3485
Qy 3407 caaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccggggtt 3466
Db 3486 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGATCCCTGGGTT 3545
Qy 3467 tgcgaccttagcagccccactctacccgctaaccaaagaaaaaggggggattctcctgggc 3526
Db 3546 TGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACGGGGACTCTGTTTGAGTGGGG 3605
Qy 3527 tcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctct 3586
Db 3606 CCCAGACCAGCAAAAGGCCTACCAAGAGATCAAGCAGGCTCTCTTAACTGCCCCTGCCCT 3665
Qy 3587 ggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaagggagtagc 3646
Db 3666 GGGATTGCCAGACTTGACTAAGCCCTTCGAACTTTTTGTTGACGAGAAGCAGGGCTACGC 3725
Qy 3647 ccgaggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctgtcaaa 3706
Db 3726 CAAAGGTGTCCTAACGCAAAAACTGGGGCCTTGGCGTCGGCCGGTGGCCTACCTGTCCAA 3785
Qy 3707 gaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggc 3766
Db 3786 AAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAGCAGCCATCGC 3845
Qy 3767 catactggtcaaggacgctgacaaattgactttgggacagaatataactgtaatagcccc 3826
Db 3846 CGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAGCCACTAGTCATTCTGGCCCC 3905
Qy 3827 ccatgcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgcccgcat 3886
Db 3906 CCATGCAGTAGAGGCACTAGTTAAGCAACCCCCTGATCGCTGGCTCTCCAACGCCCGAAT 3965
Qy 3887 gacccactatcaaagcctgcttct---cacagagagggtcactttcgctccaccagccgc 3943
Db 3966 GACCCACTACCAGGCTCTGCTTCTGGACACGGACCGAGTCCAGTTCGGACCAATAGTGGC 4025
Qy 3944 tctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgcca 4003
Db 4026 CCTAAACCCAGCTACGCTGCTCCCTCTA---CCTGAGGAGGGGCTGCAACATGACTGCCT 4082
Qy 4004 tcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgactgg 4063
Db 4083 TGACATCTTGGCTGAAGCCCACGGAACTAGACCAGATCTTACGGACCAGCCTCTCCCAGA 4142
Qy 4064 agaagtgctaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatggc 4123
Db 4143 CGCTGACCACACCTGGTACACAGATGGGAGCAGCTTCCTGCAAGAGGGGCAGCGCAAGGC 4202
Qy 4124 tggggcggcagtggtggacgggacccgcacgatctgggccagcagcctgccggaaggaac 4183
Db 4203 CGGAGCAGCAGTAACCACCGAGACCGAGGTAGTCTGGGCCAAAGCACTGCCAGCCGGGAC 4262
Qy 4184 ttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaagggaa 4243
Db 4263 ATCGGCCCAAAGAGCTGAGTTGATAGCGCTCACCCAAGCCTTAAAAATGGCAGAAGGTAA 4322
Qy 4244 atccataaacatttatacggacagcaggtatgcctttgcgactgcacacgtacacggggc 4303
Db 4323 GAAGCTGAATGTTTACACCGATAGCCGTTATGCTTTTGCCACTGCCCATATTCACGGAGA 4382
Qy 4304 catctataaacaaaggggggttgcttacctcagcagggagggaaataaagaacaaagagga 4363
Db 4383 AATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAAATCAAAAATAAGGACGA 4442
Qy 4364 aattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgtcc 4423
Db 4443 GATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGACTTAGCATAATTCATTGCCC 4502
Qy 4424 tggacatcagaaagccaaagatctcatatctagagggaaccagatggctgaccgggttgc 4483
Db 4503 GGGACATCAGAAGGGAAACCGCGCGGAGGCAAGGGGCAACAGGATGGCCGACCAAGCGGC 4562
Qy 4484 caagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaaagccccaga 4543
Db 4563 CCGAGAAGTAGCCACTAGAGAAACTCCAGAGACTTCCACACTTCTGATAGAAAATTCAGC 4622
Qy 4544 accc-agacgacagtacaccctagaagactggcaagagatataaaaaagatagaccagttct 4602
Db 4623 CCCCTATACTCATGAACATTTTCACTATACGGTGACTGACATAAAAGATCTGACTAAACT 4682
Qy 4603 ctgagactccggaggg--------gacctgctatacctcatatgggaaggaaatcctgcc 4654
Db 4683 AGGGGCCACTTATGACGATGCAAAGAAGTGTTGGGTTTATCAGGGAAAGCCTGTAATGCC 4742
Qy 4655 ccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctagg------ 4708
Db 4743 TGATCAATTCACCTTTGAACTATTAGATTTTCTTCATCAATTGACCCACCTCAGTTTCTC 4802
Qy 4709 aactaaacacctgcagcagttggtcagaacatccccttatcatgttctgaggctaccagg 4768
Db 4803 AAAAACAAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTACATGCTGAACCGGGATCG 4862
Qy 4769 agtggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttc 4828
Db 4863 AACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAGGTCAATGCCAGCAAGTC 4922

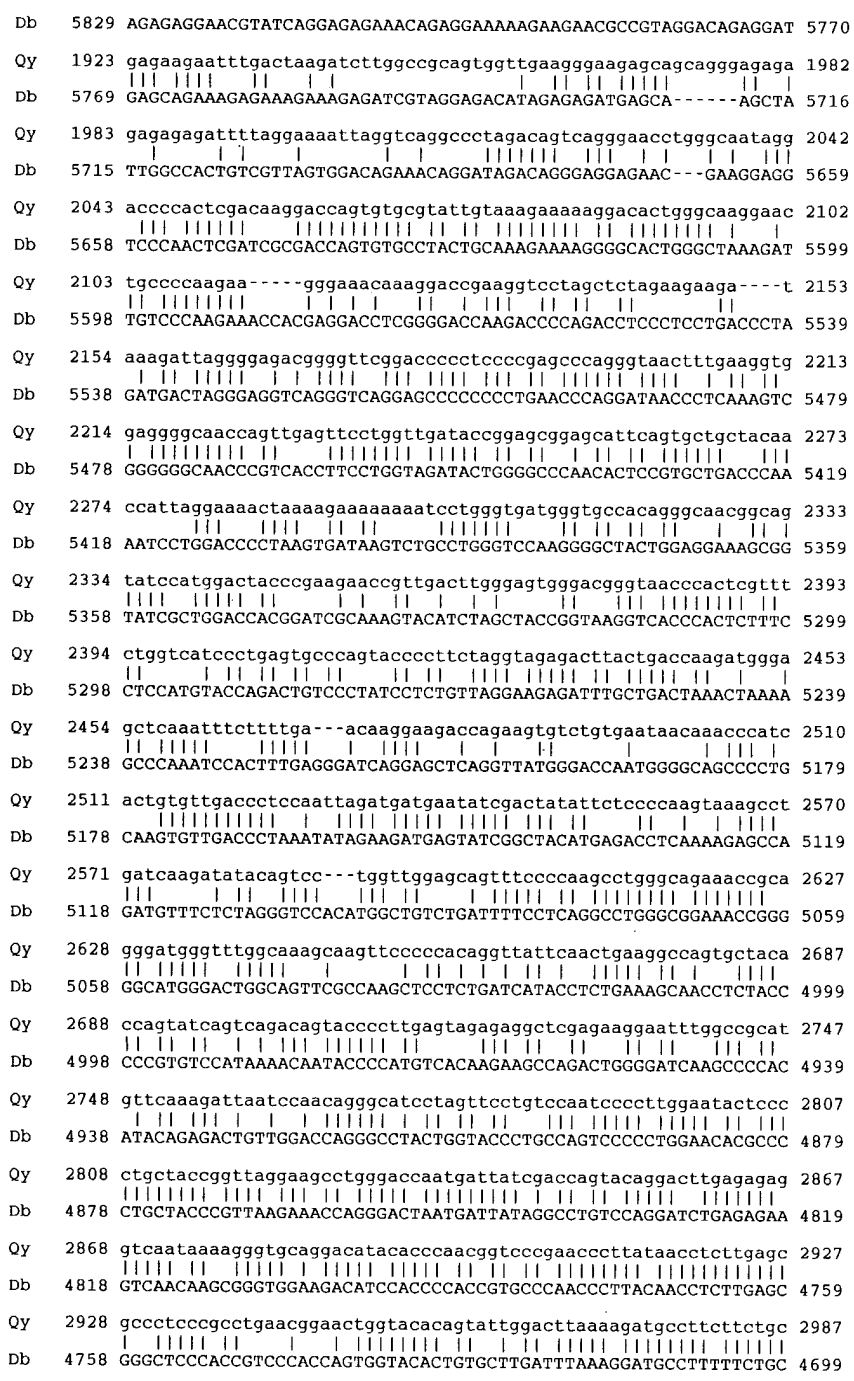
```
Db    7077 ACTCCATGCCTATCTGCCACCGTGCTTAATCGCACCACTGACTATTGCGTTCTCGTGGAA 7136

Qy    6916 atcgtcccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcgg 6975

Db    7137 TTATGGCCCAGGGTCACCTACCATCCTCCCAGTTACGTCTATAGCCAGTTTGAAAAATCC 7196

Qy    6976 tataaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggattaggg 7035

Db    7197 CATA---GACATAAAAGAGAACCAGTGTCCTTAACCTTGGCCTTATTATTAGGTGGGCTA 7253

Qy    7036 acggccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagag 7095

Db    7254 ACTATGGGTGGCATCGCCGCGGGAGTAGGGACAGGAACTACCG---CCCTGGTCGCCACC 7310

Qy    7096 aaaggacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtct 7155

Db    7311 CAGCAGTTTCAGCAGCTCCATGCTGCCGTACAAGATGATCTCAAAGAAGTCGAAAAGTCA 7370

Qy    7156 gttagcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggagg 7215

Db    7371 ATTACTAACCTAGAAAAGTCTCTTACTTCGTTGTCTGAGGTTGTACTGCAGAATCGACGA 7430

Qy    7216 ggattagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgt 7275

Db    7431 GGCCTAGACCTGTTGTTCCTAAAAGAGGGAGGACTGTGTGCTGCCCTAAAAGAAGAATGT 7490

Qy    7276 tgcttctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaag 7335

Db    7491 TGTTTCTATGCTGACCATACAGGCCTAGTAAGAGATAGTATGGCCAAATTAAGAGAGAGA 7550

Qy    7336 ttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaac 7395

Db    7551 CTCTCTCAGAGACAAAAACTATTTGAGTCGAGCCAAGGATGGTTCGAAGGATGGTTTAAC 7610

Qy    7396 aggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctc 7455

Db    7611 AGATCCCCCTGGTTTACCACGTTGATATCCACCATCATGGGGCCTCTCATTATACTCCTA 7670

Qy    7456 ctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacga 7515

Db    7671 CTAATTCTGCTTTTTGGACCCTGCATTCTTAATCGATTAGTTCAATTTGTTAAAGACAGG 7730

Qy    7516 gtgagtgcagtccagatcatggtacttaggcaacagtacca 7556

Db    7731 ATCTCAGTAGTCCAGGCTTTAGTCCTGACTCAACAATACCA 7771
```

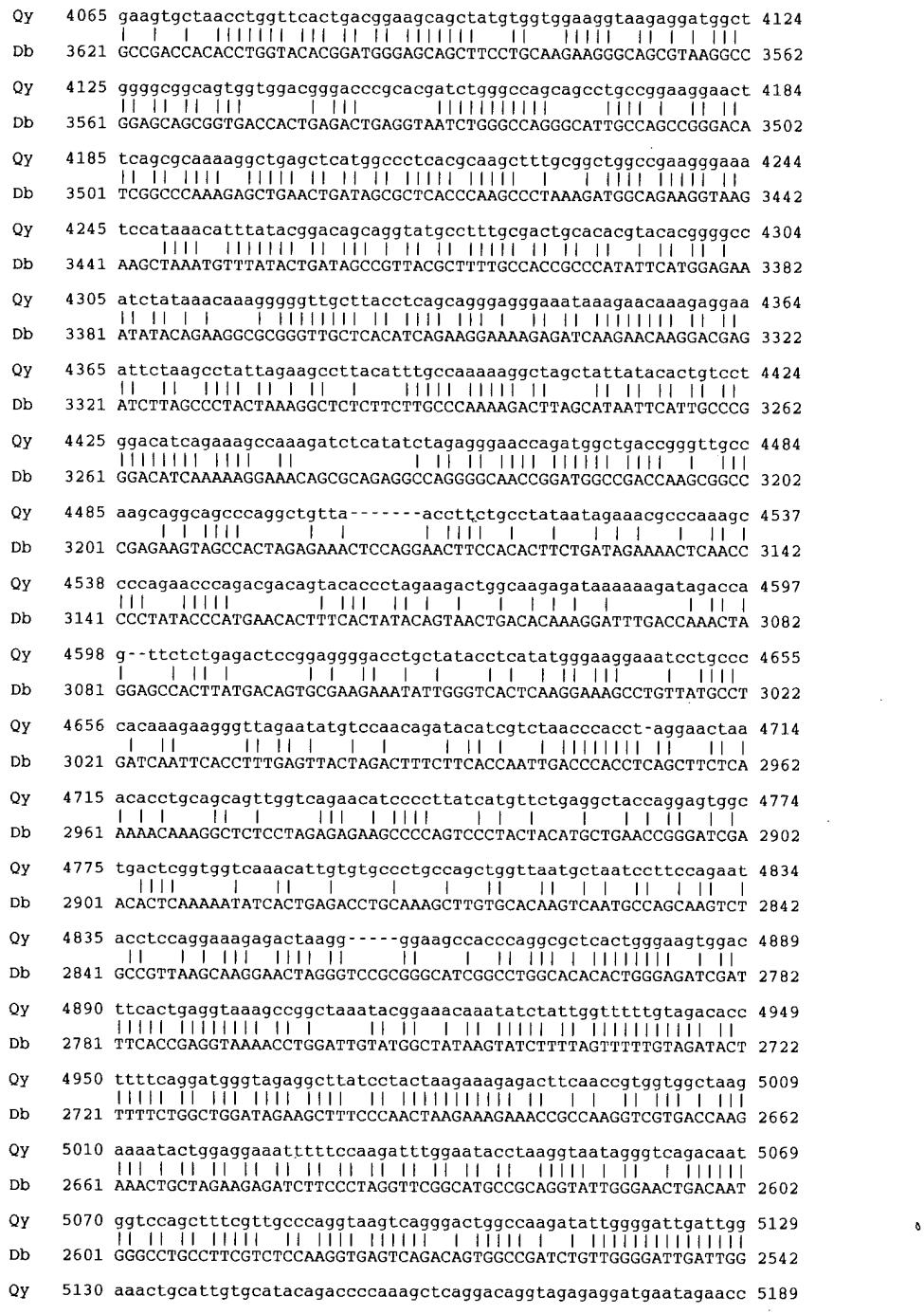
```
RESULT   9
PCT-US93-08041-8
; Sequence 8, Application PC/TUS9308041
;  GENERAL INFORMATION:
;    APPLICANT:  The Public Health Research Institute of the City of
;    APPLICANT:  New York, Inc.
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08041
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/938,100
;      FILING DATE:  28-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/_____
;      FILING DATE:  20-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-08041-8
```

```
  Query Match             21.3%;  Score 1747;  DB 5;  Length 8323;
  Best Local Similarity   55.6%;  Pred. No. 0;
  Matches 4058;  Conservative    0;  Mismatches 3005;  Indels  238;  Gaps   27;

Qy     467 tccgcggccgtccgactcttttgcctgcttgtggaa--gacgcggacgggtcgcgtgtgt 524

Db     498 TCCGCCCCCGTCTGAGTTTTTGCTTTCGGTTTGGAACCGAAGCCGCGCCGCGCGTCTTGT 557

Qy     525 ctggatctgttggtttctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagtttt 584

Db     558 CTGCTGCAGCATCGTTCTGTGTTGTCTCTGTTTGACTGTT--TTTCTGTATTTGTCTGAA 615

Qy     585 aatatgggacagacagtgactaccccccttagtttgactctcgaccattggactgaagtt 644

Db     616 AACATGGGCCAGGCTGTTACCACCCCCTTAAGTTTGACTTTAGACCACTGGAAGGATGTC 675

Qy     645 agatccagggctcataatttgtcagttcaggttaagaagggaccttggcagactttctgt 704

Db     676 GAACGGACAGCCCACAACCTGTCGGTAGAGGTTAGAAAAAGGCGCTGGGTTACATTCTGC 735

Qy     705 gcctctgaatggccaacattcgatgttggatggccatcagaggggacctttaattctgaa 764

Db     736 TCTGCAGAATGGCCAACCTTCAACGTCGGATGGCCACGAGACGGCACTTTTAACCCAGAC 795

Qy     765 attatcctggctgttaaggcaatcatttttcagactggacccggctctcatcctgatcag 824

Db     796 ATTATTACACAGGTTAAGATCAAGGTCTTCTCACCTGGCCCACATGGACATCCGGATCAG 855

Qy     825 gagccctatatccttacgtggcaagatttggcagaagatcctccgccatgggttaaacca 884

Db     856 GTCCCCTACATCGTGACCTGGGAAGCTATAGCAGTAGACCCCCCTCCCTGGGTCAGACCC 915

Qy     885 tggctaaa----------taaaccaagaaagccaggtccccgaatcctggc 925

Db     916 TTCGTGCACCCTAAACCTCCCCTCTCTCTTCCCCCTTCAGCCCCCTCTCTCCCACCTGAA 975

Qy     926 tcttggagagaaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctacc 985

Db     976 CCCCCACTCTCGACCCCGCCCCAGTCCTCCCTCTATCCGGCTCTCACTTCTCCTTTAAAC 1035

Qy     986 ccgagatcgaggagccgccgacttggccggaaccccaacctgttcccccaccccccttatc 1045

Db    1036 ACCAAACCTAGGCCTCAAGTCCTTCCTGATAGCGGAGGACCACTCATTGATCTACTCACG 1095

Qy    1046 cagcacagggtgctgtgaggggacctctgcccctcctggagctccggtggtggagggacc 1105

Db    1096 GAGGACCCTCCGCCTTACCGGGACCCAGGGCCACCCTCTCCTGACGGGAACGGCGATAGC 1155

Qy    1106 tgctgccgggactc-------ggagccggagaggcgccacc----------ccg 1142
```


Db 5523 AGTCTGGAGACCGTTGGCGGCAGCTTACCAAGAACAACTGGACCGGCCGGTAGTGCCTCA 5582
Qy 5483 tcgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagac 5542
Db 5583 CCCTTTCCGAGTCGGTGACACAGTGTGGGTCCGCAGACACCAAACTAAAAATCTAGAACC 5642
Qy 5543 tcggtggaagggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaagg 5602
Db 5643 CCGCTGGAAAGGACCTTATACCGTCCTACTGACTACCCCCACCGCTCTCAAAGTGGACGG 5702
Qy 5603 aatccccacctggatccatgcatcccacgttaagccggc---------------gccacc 5647
Db 5703 CATTGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGACACCAGGATTGAGCCACC 5762
Qy 5648 tcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcgcctccatcg 5707
Db 5763 ATCGGAATCGACATGGCGTGTTCAACGCTCTCAAAATCCCCTAAAGATAAGATTGACCCG 5822
Qy 5708 cctggttccttactctaacaataactccccaggccagtagtaaacgccttatagacagct 5767
Db 5823 CGGGACCTCCTAATCCCCTTAATTCTCTTCCTGTCTCTCAAAGGGGCCAGATCCGCAGCA 5882
Qy 5768 cgaaccccca----tagacctttatcccttacctggctgattattgaccctgatacgggt 5823
Db 5883 CCCGGCTCCAGCCCTCACCAGGTCTACAACATTACCTGGGAAGTGACCAATGGGGATCGG 5942
Qy 5824 gtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaac----- 5878
Db 5943 GAGACAGTATGGGCAATATCAGGCAACCACCCTCTGTGGACTTGGTGGCCAGTCCTCACC 6002
Qy 5879 ---------tgcatttctgcctccgattgattaaccccgctgttaaaagcacacctccca 5929
Db 6003 CCAGATTTGTGTATGTTAGCTCTCAGTGGGCCGCCCCACTGGGGGCTAGAGTATCAGGCC 6062
Qy 5930 acctagtccgtagttatgggttc----tattgctgcccaggcacagagaaagagaaatac 5985
Db 6063 CCCTATTCCTCGCCCCCGGGGCCCCCTTGTTGCTCAGGGAGCAGCGGGAACGTTGCAGGC 6122
Qy 5986 tgtggggg----------------------------ttctggggaatccttctgtagga 6016
Db 6123 TGTGCCAGAGACTGCAACGAGCCCTTGACCTCCCTCACCCCTCGGTGCAACACTGCCTGG 6182
Qy 6017 gatggagctgcgtcacctccaacgatggagactggaaatggccgatctctctccaggacc 6076
Db 6183 AACAGACTTAAGCTGGACCAGGTAACTCATAAATCAAGTGAGGGATTTTATGTCTGCCCC 6242
Qy 6077 gggtaaaattctcctttgtcaattccggcccgggcaagtacaaaatgatgaaact----- 6131
Db 6243 GGGTCACATCGCCCCCGGGAAGCCAAGTCCTGTGGGGGTCCAGACTCCTTCTACTGTGCC 6302
Qy 6132 --------------------------atataaagataagagctgctccccatcagactta 6165
Db 6303 TCTTGGGGCTGCGAGACAACCGGTAGAGTATACTGGAAGCCCTCCTCTTCTTGGGACTAC 6362
Qy 6166 gattatctaaagataagtttcactgaaaggaaaacaggaaaatattcaaaagtggataaa 6225
Db 6363 ATCACAGTAGACAACAATCTCACCTCTAACCAGGCTGTTCAGGTATGCAAAGACAATAAG 6422
Qy 6226 tggtatgagctggggaatagttttttattatatggcgggggagcagggtccactttaacc 6285
Db 6423 TGGTGCAATCCCTTGGCTATCCGGTTTACAAACGCCGGGAAACAGGTCACCTCATGGACA 6482
Qy 6286 a---------------ttcgccttaggatagagacggggacagaaccccctgtggcaatg 6330
Db 6483 ACTGGACACTATTGGGGTCTACGTCTTTATGTCTCTGGACAGGACCCAGGGCTTACTTTC 6542
Qy 6331 ggacccgataaagtactggctgaacaggggccccggccctggagccaccgcataacttg 6390
Db 6543 GGGATCCGACTCAGTTATCAAAATCTAGGACCTCGGATCCCAATAGGACCAAACCCCGTC 6602
Qy 6391 ccggtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtacc 6450
Db 6603 CTGGCAGACCAACTTTCGTTCCCGCTACCTAATCCCCTACCCAAACCTGCCAAGTCTCCC 6662
Qy 6451 actggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaag------ 6504
Db 6663 CCCGCCTCTAGTTCGACTCCCACATTGATTTCCCCGTCCCCCACTCCCACTCAGCCCCCG 6722
Qy 6505 ---------acaggacagagactcttcagtctcatccagggagctttccaagccatcaac 6555
Db 6723 CCAGCAGGAACGGGAGACAGATTACTAAATCTAGTACAGGGAGCTTACCAGGCACTCAAC 6782
Qy 6556 tccaccgaccctgatgccacttcttcttgttggctttgtctatcctcagggcctccttat 6615
Db 6783 CTTACCAACCCTGATAAAACTCAAGAGTGCTGGTTATGCCTAGTGTCTGGACCCCCCTAT 6842
Qy 6616 tatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgt 6675
Db 6843 TACGAGGGGGTTGCCGTCCTAGGTACTTATTCCAACCATACCTCTGCCCCAGCTAACTGC 6902
Qy 6676 acatgggggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgcata 6735
Db 6903 TCCGTGGCCTCCCAACACAAGCTGACCCTGTCCGAAGTGACTGGACGGGGACTCTGCATA 6962
Qy 6736 ggaaaagctcccccatcccaccaacacctttgctatagtactgtggtttatgagcaggcc 6795
Db 6963 GGAACAGTCCCAAAAACTCACCAGGCCCTGTGCAACACTACCCT------TAAGGCAGGC 7016
Qy 6796 tcagaaaatcagtatttagtacctggttataacaggtggtgggcatgcaatactgggtta 6855
Db 7017 AAAGGGTCTTACTATCTAGTTGCCCCCACAGGAACTATGTGGGCATGTAACACTGGACTC 7076
Qy 6856 acccccctgtgtttccacctcagtcttcaaccaatccaaagatttctgtgtcatggtccaa 6915
Db 7077 ACTCCATGCCTATCTGCCACCGTGCTTAATCGCACCACTGACTATTGCGTTCTCGTGGAA 7136
Qy 6916 atcgtcccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcgg 6975
Db 7137 TTATGGCCCAGGGTCACCTACCATCCTCCCAGTTACGTCTATAGCCAGTTTGAAAAATCC 7196
Qy 6976 tataaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggattaggg 7035
Db 7197 CATA---GACATAAAAGAGAACCAGTGTCCTTAACCTTGGCCTTATTATTAGGTGGGCTA 7253
Qy 7036 acggccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagag 7095
Db 7254 ACTATGGGTGGCATCGCCGCGGGAGTAGGGACAGGAACTACCG---CCCTGGTCGCCACC 7310
Qy 7096 aaaggacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtct 7155
Db 7311 CAGCAGTTTCAGCAGCTCCATGCTGCCGTACAAGATGATCTCAAAGAAGTCGAAAAGTCA 7370
Qy 7156 gttagcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggagg 7215
Db 7371 ATTACTAACCTAGAAAAGTCTCTTACTTCGTTGTCTGAGGTTGTACTGCAGAATCGACGA 7430
Qy 7216 ggattagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgt 7275
Db 7431 GGCCTAGACCTGTTGTTCCTAAAAGAGGGAGGACTGTGTGCTGCCCTAAAAGAAGAATGT 7490
Qy 7276 tgcttctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaag 7335
Db 7491 TGTTTCTATGCTGACCATACAGGCCTAGTAAGAGATAGTATGGCCAAATTAAGAGAGAGA 7550
Qy 7336 ttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaac 7395
Db 7551 CTCTCTCAGAGACAAAAACTATTTGAGTCGAGCCAAGGATGGTTCGAAGGATGGTTTAAC 7610
Qy 7396 aggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctc 7455
Db 7611 AGATCCCCCTGGTTTACCACGTTGATATCCACCATCATGGGGCCTCTCATTATACTCCTA 7670
Qy 7456 ctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacga 7515
Db 7671 CTAATTCTGCTTTTTGGACCCTGCATTCTTAATCGATTAGTTCAATTTGTTAAAGACAGG 7730

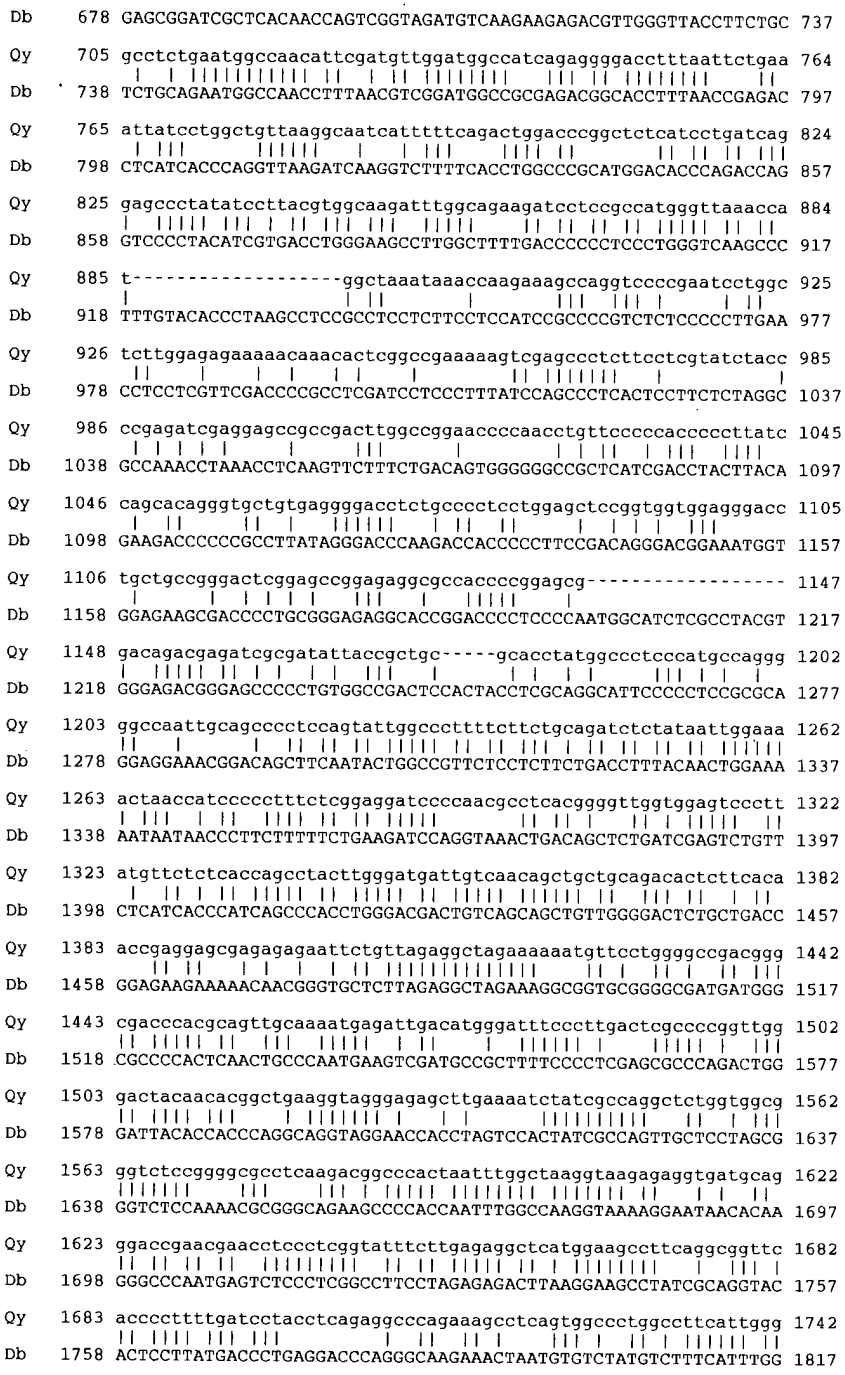
```
Qy   7516 gtgagtgcagtccagatcatggtacttaggcaacagtacca 7556
Db   7731 ATCTCAGTAGTCCAGGCTTTAGTCCTGACTCAACAATACCA 7771


RESULT  10
US-08-258-420-13/c
; Sequence 13, Application US/08258420
; Patent No. 5710037
;  GENERAL INFORMATION:
;    APPLICANT:  Nienhuis, Arthur W.
;    APPLICANT:  Vanin, Elio F.
;    TITLE OF INVENTION:  No. 5710037el Retroviral Envelope and LTR and Retroviral Vectc
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan,Cecchi, Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch diskette
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,420
;      FILING DATE:  10-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Olstein, Elliot M.
;      REGISTRATION NUMBER:  24,025
;      REFERENCE/DOCKET NUMBER:  271010-208
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8202 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  viral genome
US-08-258-420-13


  Query Match             21.2%;  Score 1738.8;  DB 1;  Length 8202;
  Best Local Similarity   55.6%;  Pred. No. 0;
  Matches 3932;  Conservative    0;  Mismatches 3012;  Indels  126;  Gaps   26;

Qy    585 aatatgggacagacagtgactacccccttagtttgactctcgaccattggactgaagtt 644
Db   7149 AAAATGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAAAGATGTC 7090

Qy    645 agatccagggctcataatttgtcagttcaggttaagaagggaccttggcagactttctgt 704
Db   7089 GAGCGGATCGCTCACAACCAGTCGGTAGATGTCAAGAAGAGACGTTGGGTTACCTTCTGC 7030

Qy    705 gcctctgaatggccaacattcgatgttggatggccatcagaggggacctttaattctgaa 764
Db   7029 TCTGCAGAATGGCCAACCTTTAACATCGGATGGCCGCGAGACGGCACCTTTAACCGAGAC 6970

Qy    765 attatcctggctgttaaggcaatcatttttcagactggacccggctctcatcctgatcag 824
Db   6969 CTCATCACCCAGGTTAAGATCAAGGTCTTTTCACCTGGCCCGCATGGACACCCAGACCAG 6910

Qy    825 gagccctatatccttacgtggcaagatttggcagaagatcctccgccatgggttaaacca 884
Db   6909 GTCCCCTACATCGTGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGCCC 6850

Qy    885 t------------------ggctaaataaaccaagaaagccaggtccccgaatcctggc 925
Db   6849 TTTGTACACCCTAAGCCTCCGCCTCCTCTTCCTCCATCCGCCCCGTCTCTCCCCCTTGAA 6790

Qy    926 tcttggagagaaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctacc 985
Db   6789 CCTCCTCGTTCGACCCCGCCTCGCTCCTCCCTTTATCCAGCCCTCACTCCTTCTCTAGGC 6730

Qy    986 ccgagatcgaggagccgccgacttggccggaaccccaacctgttcccccaccccccttatc 1045
Db   6729 GCCAAACCTAAACCTCAAGTTCTTTCTGACAGTGGGGGGCCGCTCATCGACCTACTTACA 6670

Qy   1046 cagcacagggtgctgtgaggggacctctgcccctcctggagctccggtggtggagggacc 1105
Db   6669 GAAGACCCCCCGCCTTATAGGGACCCAAGACCACCCCCTTCCGACAGGGACGGAAATGGT 6610

Qy   1106 tgctgccgggactcggagccggagaggcgccaccccggagc------------------g 1147
Db   6609 GGAGAAGCGACCCCTGCGGGAGAGGCACCGGACCCCTCCCCAATGGCATCTCGCCTACGT 6550

Qy   1148 gacagacgagatcgcgatattaccgctgc-----gcacctatggccctcccatgccaggg 1202
Db   6549 GGGAGACGGGAGCCCCCTGTGGCCGACTCCACTACCTCGCAGGCATTCCCCCTCCGCGCA 6490

Qy   1203 ggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattggaaa 1262
Db   6489 GGAGGAAACGGACAGCTTCAATACTGGCCGTTCTCCTCTTCTGACCTTTACAACTGGAAA 6430

Qy   1263 actaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtccctt 1322
Db   6429 AATAATAACCCTTCTTTTTCTGAAGATCCAGGTAAACTGACAGCTCTGATCGAGTCTGTT 6370

Qy   1323 atgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttcaca 1382
Db   6369 CTCATCACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTGGGGACTCTGCTGACC 6310

Qy   1383 accgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgacggg 1442
Db   6309 GGAGAAGAAAAACAACGGGTGCTCTTAGAGGCTAGAAAGGCGGTGCGGGGCGATGATGGG 6250

Qy   1443 cgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggttgg 1502
Db   6249 CGCCCCACTCAACTGCCCAATGAAGTCGATGCCGCTTTTCCCCTCGAGCGCCCAGACTGG 6190

Qy   1503 gactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcg 1562
Db   6189 GATTACACCACCCAGGCAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCG 6130

Qy   1563 ggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcag 1622
Db   6129 GGTCTCCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGAATAACACAA 6070

Qy   1623 ggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcggttc 1682
Db   6069 GGGCCCAATGAGTCTCCCTCGGCCTTCCTAGAGAGACTTAAGGAAGCCTATCGCAGGTAC 6010

Qy   1683 acccctttgatcctacctcagaggcccagaaagcctcagtggccctggccttcattggg 1742
Db   6009 ACTCCTTATGACCCTGAGGACCCAGGGCAAGAAACTAATGTGTCTATGTCTTTCATTTGG 5950

Qy   1743 cagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggctgag 1802
Db   5949 CAGTCTGCCCCAGACATTGGGAGAAAGTTAGAGAGGTTAGAAGATTTAAAAAACAAGACG 5890

Qy   1803 ttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagaggag 1862
Db   5889 CTTGGAGATTTGGTTAGAGAGGCAGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAA 5830

Qy   1863 aaggaacagagaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacggcaa 1922
```

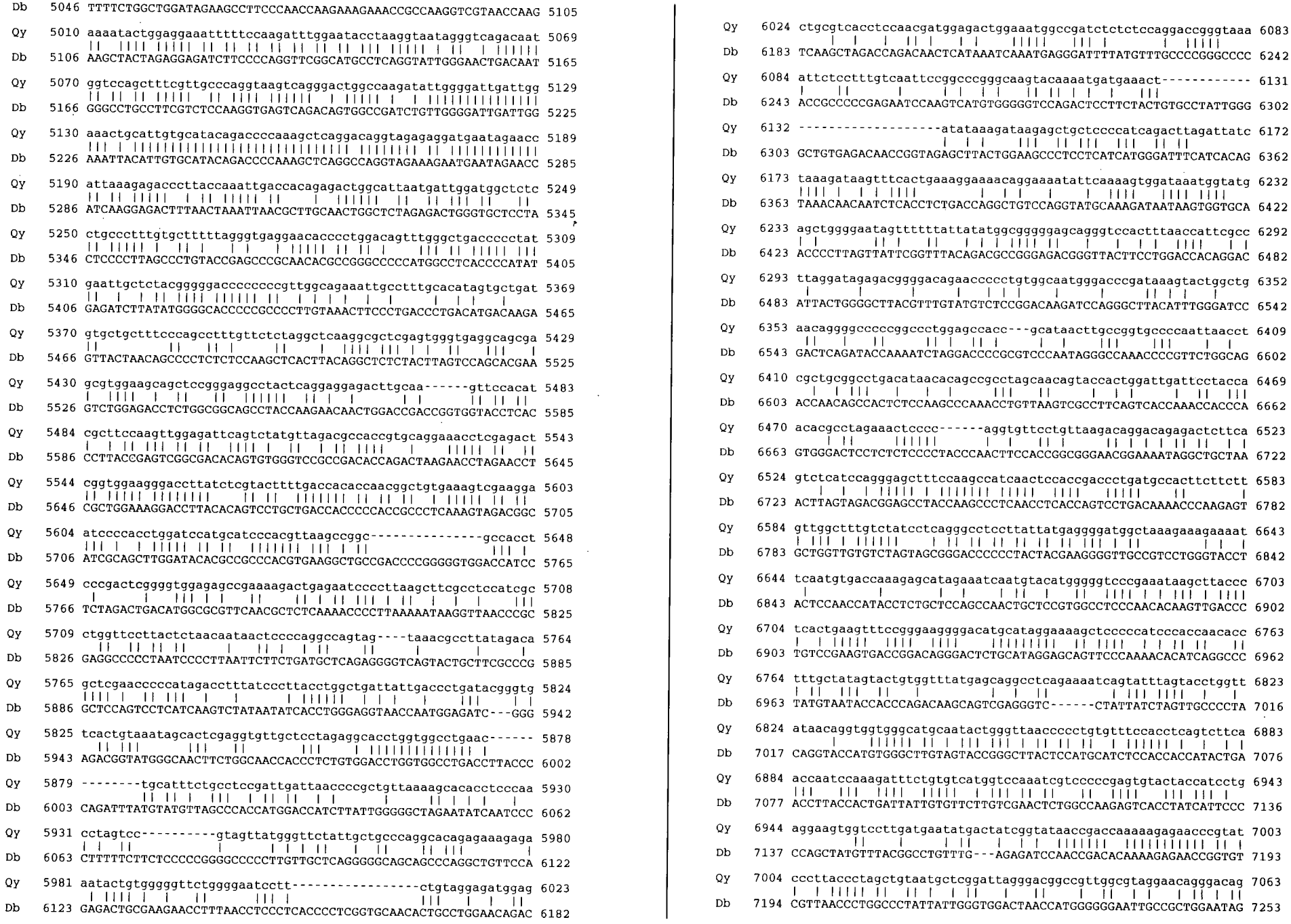

Db 5829 AGAGAGGAACGTATCAGGAGAGAAACAGAGGAAAAAGAAGAACGCCGTAGGACAGAGGAT 5770
Qy 1923 gagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggagaga 1982
Db 5769 GAGCAGAAAGAGAAAGAAAGAGATCGTAGGAGACATAGAGAGATGAGCA------AGCTA 5716
Qy 1983 gagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaatagg 2042
Db 5715 TTGGCCACTGTCGTTAGTGGACAGAAACAGGATAGACAGGGAGGAGAAC---GAAGGAGG 5659
Qy 2043 accccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaaggaac 2102
Db 5658 TCCCAACTCGATCGCGACCAGTGTGCCTACTGCAAAGAAAAGGGGCACTGGGCTAAAGAT 5599
Qy 2103 tgccccaagaa-----gggaaacaaaggaccgaaggtcctagctctagaagaaga----t 2153
Db 5598 TGTCCCAAGAAACCACGAGGACCTCGGGGACCAAGACCCCAGACCTCCCTCCTGACCCTA 5539
Qy 2154 aaagattaggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtg 2213
Db 5538 GATGACTAGGGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTC 5479
Qy 2214 gaggggcaaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaa 2273
Db 5478 GGGGGGCAACCCGTCACCTTCCTGGTAGATACTGGGGCCCAACACTCCGTGCTGACCCAA 5419
Qy 2274 ccattaggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcag 2333
Db 5418 AATCCTGGACCCCTAAGTGATAAGTCTGCCTGGGTCCAAGGGGCTACTGGAGGAAAGCGG 5359
Qy 2334 tatccatggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgttt 2393
Db 5358 TATCGCTGGACCACGGATCGCAAAGTACATCTAGCTACCGGTAAGGTCACCCACTCTTTC 5299
Qy 2394 ctggtcatccctgagtgcccagtacccctttctaggtagagacttactgaccaagatggga 2453
Db 5298 CTCCATGTACCAGACTGTCCCTATCCTCTGTTAGGAAGAGATTTGCTGACTAAACTAAAA 5239
Qy 2454 gctcaaatttcttttga---acaaggaagaccagaagtgtctgtgaataacaaacccatc 2510
Db 5238 GCCCAAATCCACTTTGAGGGATCAGGAGCTCAGGTTATGGGACCAATGGGGCAGCCCCTG 5179
Qy 2511 actgtgttgaccctccaattagatgatgaatatcgactatattctccccaagtaaagcct 2570
Db 5178 CAAGTGTTGACCCTAAATATAGAAGATGAGTATCGGCTACATGAGACCTCAAAAGAGCCA 5119
Qy 2571 gatcaagatatacagtcc---tggttggagcagtttccccaagcctgggcagaaaccgca 2627
Db 5118 GATGTTTCTCTAGGGTCCACATGGCTGTCTGATTTTCCTCAGGCCTGGGCGGAAACCGGG 5059
Qy 2628 gggatgggtttggcaaagcaagttcccccacaggttattcaactgaaggccagtgctaca 2687
Db 5058 GGCATGGGACTGGCAGTTCGCCAAGCTCCTCTGATCATACCTCTGAAAGCAACCTCTACC 4999
Qy 2688 ccagtatcagtcagacagtacccctgagtagagaggctcgagaaggaatttggccgcat 2747
Db 4998 CCCGTGTCCATAAAACAATACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCAC 4939
Qy 2748 gttcaaagattaatccaacagggcatcctagttcctgtccaatccccttggaatactccc 2807
Db 4938 ATACAGAGACTGTTGGACCAGGGCCTACTGGTACCCTGCCAGTCCCCCTGGAACACGCCC 4879
Qy 2808 ctgctaccggttaggaagcctgggaccaatgattatcgaccagtacaggacttgagagag 2867
Db 4878 CTGCTACCCGTTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGGATCTGAGAGAA 4819
Qy 2868 gtcaataaaagggtgcaggacatacacccaacggtcccgaaccccttataacctcttgagc 2927
Db 4818 GTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCCAACCCTTACAACCTCTTGAGC 4759
Qy 2928 gccctcccgcctgaacggaactggtacacagtattggacttaaaagatgccttcttctgc 2987
Db 4758 GGGCTCCCACCGTCCCACCAGTGGTACACTGTGCTTGATTTAAAGGATGCCTTTTTCTGC 4699
Qy 2988 ctgagattacaccccactagccaaccacttttgccttcgaatggagagatccaggtacg 3047
Db 4698 CTGAGACTCCACCCCACCAGTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATG 4639
Qy 3048 ggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactccccgacc 3107
Db 4638 GGAATCTCAGGACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCACC 4579
Qy 3108 atctttgacgaagccctacacagggacctggccaacttcaggatccaacaccctcaggtg 3167
Db 4578 CTGTTTGATGAGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGCACCCAGACTTG 4519
Qy 3168 accctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggactgctta 3227
Db 4518 ATCCTGCTACAGTACGTGGATGACTTACTGCTGGCCGCCACTTCTGAGCTAGACTGCCAA 4459
Qy 3228 gaaggtacgaaggcactactgctggaattgtctgacctaggctacagagcctctgctaag 3287
Db 4458 CAAGGTACTCGGGCCCTGTTACAAACCCTAGGGAACCTCGGGTATCGGGCCTCGGCCAAG 4399
Qy 3288 aaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggcgggcag 3347
Db 4398 AAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAG 4339
Qy 3348 cgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaaccacagcc 3407
Db 4338 AGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGCGAGCCTACTCCGAAGACCCCT 4279
Qy 3408 aaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccggggttt 3467
Db 4278 CGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGATCCCTGGGTTT 4219
Qy 3468 gcgaccttagcagccccactctacccgctaaccaaagaaaaaggggggattctcctgggct 3527
Db 4218 GCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACGGGGACTCTGTTTAATTGGGGC 4159
Qy 3528 cctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctg 3587
Db 4158 CCAGACCAACAAAAGGCCTATCAAGAAATCAAGCAAGCTCTTCTAACTGCCCCAGCCCTG 4099
Qy 3588 gccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaagggagtagcc 3647
Db 4098 GGGTTGCCAGATTTGACTAAGCCCTTTGAACTCTTTGTCGACGAGAAGCAGGGCTACGCC 4039
Qy 3648 cgaggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctgtcaaag 3707
Db 4038 AAAGGCGTCCTAACGCAAAAGCTGGGACCTTGGCGTCGGCCGGTGGCCTACCTGTCTAAA 3979
Qy 3708 aagcttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggcc 3767
Db 3978 AAGCTAGACCCAGTGGCAGCTGGCTGGCCCCCCTGCCTACGGATGGTGGCAGCCATTGCA 3919
Qy 3768 atactggtcaaggacgctgacaaattgactttgggacagaatataactgtaatagccccc 3827
Db 3918 GTTCTGACAAAAGATGCTGGCAAGCTCACTATGGGACAGCCGTTGGTCATTCTGGCCCCC 3859
Qy 3828 catgcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgcccgcatg 3887
Db 3858 CATGCCGTAGAGGCACTAGTTAAGCAACCCCCTGATCGCTGGCTCTCCAATGCCCGGATG 3799
Qy 3888 acccactatcaaagcctgcttct---cacagagagggtcactttcgctccaccagccgct 3944
Db 3798 ACCCATTACCAAGCCCTGCTCCTGGACACGGACCGGGTCCAGTTCGGGCCAGTAGTGGCC 3739
Qy 3945 ctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgccat 4004
Db 3738 CTAAATCCAGCTACGCTGCTCCCT---CTGCCTGAGGAGGGGCTGCAACATGACTGCCTT 3682
Qy 4005 caactattgattgaggagactggggtccgcaaggaccttacagacataccgctgactgga 4064
Db 3681 GACATCTTGGCTGAAGCCCACGGAACTAGATCAGATCTTACGGACCAGCCCCTCCCAGAC 3622



```
Db   1464 CCACACCGGGGGGTCGATGCAATCCCCTAGTCCTAGAATTCACTGACGCGGGTAAAAAGG 1405

Qy   6248 ttttattatatggcgggggagcagggtccactttaaccattcgccttaggatagagacgg 6307
          |   |||||     || | ||   |||   ||  | |   |     |  ||
Db   1404 CCAGCTGGGATGGCCCCAAAGTATGG-GGACTAAGACTGTACCAATCCACAGGGATCGAC 1346

Qy   6308 ggacagaacccccctgtggcaatgggacccgataaagtactggctgaacaggggccccggg 6367
          || ||   | |          || | |      |      |  | |   ||||
Db   1345 CCGGTGACCCGGTTCTCTTTGACCCGCCAGGTCCTCAATATAGGGCCCCGCATCCCCATT 1286

Qy   6368 ccctggagccaccgcataacttgcc-ggtgccccaattaacctcgctgcggcctgacata 6426
          |  | |   || ||| |||   | ||||  |       |  ||| |        |
Db   1285 GGGCCTAATCCCGTGATCACTGGCCAACTACCCCCCTCCCGACCCGTGCAGATCAGGCTC 1226

Qy   6427 acacagccgcctagcaacagtaccactggattgattcctaccaacacgcctagaaactcc 6486
           |   ||| |||   |    | | || ||        |||    | |   |    | ||
Db   1225 CCCAGGCCTCCTCAGACTCCTCCTACAGGCGCAGCCTCTATGGTCCCTGGGACTGCCCCA 1166

Qy   6487 ccaggtgttcctgttaagacaggacagagactcttcagtctcatccagggagctttccaa 6546
          ||   |   |    |  ||| ||| | || ||  | |  ||  |   | ||||| | ||||
Db   1165 CCGTCTCAACAACCTGGGACGGGAGACAGGCTGCTAAACCTGGTAGATGGAGCATACCAA 1106

Qy   6547 gccatcaactccaccgaccctgatgccacttcttcttgttggctttgtctatcctcaggg 6606
          ||  |||||  ||||   |||||    ||       || ||| | |||||    || ||
Db   1105 GCACTCAACCTCACCAGTCCTGACAAAACCCAAGAGTGCTGGTTGTGTCTGGTATCGGGA 1046

Qy   6607 cctccttattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagcataga 6666
          || || || || || ||| | ||       | | |  | |          |        |
Db   1045 CCCCCCTACTACGAAGGGGTTGCCGTCCTAGGTACTTACTCCAACCATACCTCTGCCCCA 986

Qy   6667 aatcaatgtacatgggggtcccgaaataagcttaccctcactgaagtttccgggaagggg 6726
            | | ||  |   ||  |||| | | ||||| |||||  | |||||  ||||  ||||
Db    985 GCTAACTGCTCCGCGGCCTCCCAACACAAGCTGACCCTGTCCGAAGTAACCGGACAGGGA 926

Qy   6727 acatgcataggaaaagctcccccatcccaccaacaccttgctatagtactgtggtttat 6786
             ||| |||||  || ||||  | |||| ||   ||| ||  |||  ||        |
Db    925 CTCTGCGTAGGAGCAGTTCCCAAAACCCATCAGGCCCTGTGTAATACCACCC------AA 872

Qy   6787 gagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggcatgcaat 6846
           ||  |  |  |      | ||| | |   ||         |    ||||| |||||
Db    871 AAGACGAGCGACGGGTCCTACTATCTGGCTGCTCCCGCCGGGACCATTTGGGCTTGCAAC 812

Qy   6847 actgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttctgtgtc 6906
          || ||| | || |||||  | || ||  | ||  |||| | | ||| |||||  |||||
Db    811 ACCGGGCTCACTCCCTGCCTATCTACTACTGTACTCAATCTAACCACAGATTATTGTGTA 752

Qy   6907 atggtccaaatcgtccccgagtgtactaccatcctgaggaagtggtccttgatgaatat 6966
           | ||  || ||   ||| ||||    |||||  |    ||      |   || | |   |
Db    751 TTAGTTGAACTCTGGCCCAGAGTAATTTACCACTCCCCCGATTATATGTATGGTCAGCTT 692

Qy   6967 gactatcggtataaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctc 7026
          ||  | ||   || |  |   |||||||| || |||||  | ||||| ||  |  | ||
Db    691 GAACAGCG---TACCAAATATAAAAGAGAGCCAGTATCATTGACCCTGGCCCTTCTACTA 635

Qy   7027 ggattagggacggccgttggcgtaggaacagggacagctgccctgatcacaggaccacag 7086
          |||  |  ||   |  ||  | | | | || | ||   |  || ||| |    |
Db    634 GGAGGATTAACCATGGGAGGGATTGCAGCTGGAATAGGGACGGGGACCACTGCCTTAATT 575

Qy   7087 cagctagagaaaggacttggtgagctacatgcggccatgacagaagatctccgagcctta 7146
           |     || |     |||   |||| ||||| || ||      ||| |||   |   |
Db    574 AAAACCCAGCA---GTTTGAGCAGCTTCATGCCGCTATCCAGACAGACCTCAACGAAGTC 518

Qy   7147 aaggagtctgttagcaacctagaagagtccctgacttctttgtctgaagtggttctacag 7206
           |  ||||  ||| |||||||||| |||| ||||| || |||||||||| || ||||||
Db    517 GAAAAGTCAATTACCAACCTAGAAAAGTCACTGACCTCGTTGTCTGAAGTAGTCCTACAG 458

Qy   7207 aaccggaggggattagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaa 7266
          ||||| || ||  ||||| |||| || ||||  || || ||  | || |||||| |||||
Db    457 AACCGCAGAGGCCTAGATTTGCTATTCCTAAAGGAGGGAGGTCTCTGCGCAGCCCTAAAA 398

Qy   7267 gaagaatgttgcttctatgtagatcactcaggagccatcagagactccatgaacaagctt 7326
          ||||||||||| || |||| ||| ||| | ||     | ||||||  ||||  |||  |
Db    397 GAAGAATGTTGTTTTTATGCAGACCACACGGGGCTAGTGAGAGACAGCATGGCCAAATTA 338
```

```
Qy   7327 agaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaagga 7386
          ||| ||| | |   |   | | | ||    |   ||| | | ||| || ||||| |||||
Db    337 AGAGAAAGGCTTAATCAGAGACAAAAACTATTTGAGACAGGCCAAGGATGGTTCGAAGGG 278

Qy   7387 tggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggcccctagta 7446
             ||| || || || || ||| | |||||| |  | ||  |  | | ||| || ||| ||
Db    277 CTGTTTAATAGATCCCCCTGGTTTACCACCTTAATCTCCACCATCATGGGACCTCTAATA 218

Qy   7447 gtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgcctttgtt 7506
          || ||  | ||| |  |     |||| ||||||  |  | ||| | || ||    ||||||
Db    217 GTACTCTTACTGATCTTACTCTTTGGACCTTGCATTCTCAATCGATTAGTCCAATTTGTT 158

Qy   7507 agagaacgagtgagtgcagtccagatcatggtacttaggcaacagtacca 7556
          | |||  |  |    |  ||||||    | ||  | |  ||||| || ||
Db    157 AAAGACAGGATATCAGTGGTCCAGGCTCTAGTTTTGACTCAACAATATCA 108


RESULT  11
US-08-850-961-1
; Sequence 1, Application US/08850961
; Patent No. 6013517
;  GENERAL INFORMATION:
;    APPLICANT:  Respess, James G.
;    APPLICANT:  De Polo, Nicholas J.
;    APPLICANT:  Chada, Sunil
;    APPLICANT:  Sauter, Sybille
;    APPLICANT:  Bodner, Mordechai
;    APPLICANT:  Driver, David A.
;    TITLE OF INVENTION:  CROSSLESS RETROVIRAL VECTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation, Intellectual Property - R440
;      STREET:  P.O. Box 8097
;      CITY:  Emeryville
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/850,961
;      FILING DATE:  05-MAY-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kruse, No. 6013517man J.
;      REGISTRATION NUMBER:  35,235
;      REFERENCE/DOCKET NUMBER:  930049.424C4 / 1147.005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-3520
;      TELEFAX:  (510) 655-3542
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8332 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-850-961-1


  Query Match             20.6%;  Score 1687.4;  DB 3;  Length 8332;
  Best Local Similarity   55.4%;  Pred. No. 0;
  Matches 3964;  Conservative    0;  Mismatches 2981;  Indels  208;  Gaps   28;

Qy    585 aatatgggacagacagtgactacccccttagtttgactctcgaccattggactgaagtt 644
          |||||||| ||||| || || || ||| | ||||||||  | |  || ||||  || ||
Db    618 AATATGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAAAGATGTC 677

Qy    645 agatccagggctcataatttgtcagttcaggttaagaagggaccttggcagactttctgt 704
                |  ||||| ||   ||| ||  | || ||||||| ||| ||||   || |||||
```

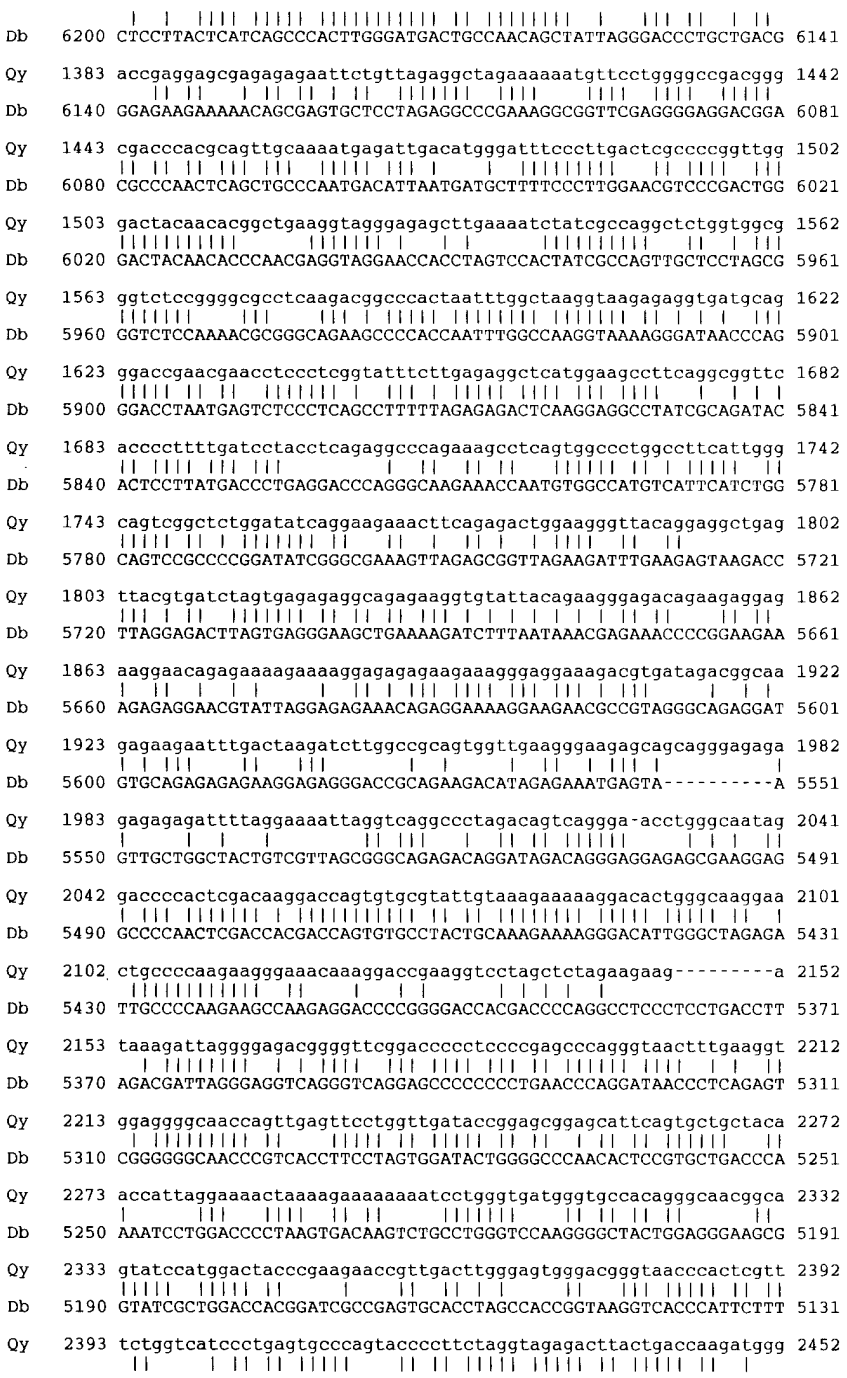

Db 678 GAGCGGATCGCTCACAACCAGTCGGTAGATGTCAAGAAGAGACGTTGGGTTACCTTCTGC 737
Qy 705 gcctctgaatggccaacattcgatgttggatggccatcagaggggacctttaattctgaa 764
Db 738 TCTGCAGAATGGCCAACCTTTAACGTCGGATGGCCGCGAGACGGCACCTTTAACCGAGAC 797
Qy 765 attatcctggctgttaaggcaatcatttttcagactggacccggctctcatcctgatcag 824
Db 798 CTCATCACCCAGGTTAAGATCAAGGTCTTTTCACCTGGCCCGCATGGACACCCAGACCAG 857
Qy 825 gagccctatatccttacgtggcaagatttggcagaagatcctccgccatgggttaaacca 884
Db 858 GTCCCCTACATCGTGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGCCC 917
Qy 885 t------------------ggctaaataaaccaagaaagccaggtccccgaatcctggc 925
Db 918 TTTGTACACCCTAAGCCTCCGCCTCCTCTTCCTCCATCCGCCCCGTCTCTCCCCCTTGAA 977
Qy 926 tcttggagagaaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctacc 985
Db 978 CCTCCTCGTTCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTCACTCCTTCTCTAGGC 1037
Qy 986 ccgagatcgaggagccgccgacttggccggaaccccaacctgttcccccaccccctatc 1045
Db 1038 GCCAAACCTAAACCTCAAGTTCTTTCTGACAGTGGGGGGCCGCTCATCGACCTACTTACA 1097
Qy 1046 cagcacagggtgctgtgaggggacctctgcccctcctggagctccggtggtggagggacc 1105
Db 1098 GAAGACCCCCCGCCTTATAGGGACCCAAGACCACCCCCTTCCGACAGGGACGGAAATGGT 1157
Qy 1106 tgctgccgggactcggagccggagaggcgccaccccggagcg------------------ 1147
Db 1158 GGAGAAGCGACCCCTGCGGGAGAGGCACCGGACCCCTCCCCAATGGCATCTCGCCTACGT 1217
Qy 1148 gacagacgagatcgcgatattaccgctgc-----gcacctatggccctcccatgccaggg 1202
Db 1218 GGGAGACGGGAGCCCCCTGTGGCCGACTCCACTACCTCGCAGGCATTCCCCCTCCGCGCA 1277
Qy 1203 ggccaattgcagcccctccagtattggccctttttcttctgcagatctctataattggaaa 1262
Db 1278 GGAGGAAACGGACAGCTTCAATACTGGCCGTTCTCCTCTTCTGACCTTTACAACTGGAAA 1337
Qy 1263 actaaccatccccctttctcggaggatccccaacgcctcacggggttggtggagtccctt 1322
Db 1338 AATAATAACCCTTCTTTTTCTGAAGATCCAGGTAAACTGACAGCTCTGATCGAGTCTGTT 1397
Qy 1323 atgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttcaca 1382
Db 1398 CTCATCACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTGGGGACTCTGCTGACC 1457
Qy 1383 accgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgacggg 1442
Db 1458 GGAGAAGAAAAACAACGGGTGCTCTTAGAGGCTAGAAAGGCGGTGCGGGGCGATGATGGG 1517
Qy 1443 cgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggttgg 1502
Db 1518 CGCCCCACTCAACTGCCCAATGAAGTCGATGCCGCTTTTCCCCTCGAGCGCCCAGACTGG 1577
Qy 1503 gactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcg 1562
Db 1578 GATTACACCACCCAGGCAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCG 1637
Qy 1563 ggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcag 1622
Db 1638 GGTCTCCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGAATAACACAA 1697
Qy 1623 ggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcggttc 1682
Db 1698 GGGCCCAATGAGTCTCCCTCGGCCTTCCTAGAGAGACTTAAGGAAGCCTATCGCAGGTAC 1757
Qy 1683 acccctttgatcctacctcagaggcccagaaagcctcagtggccctggccttcattggg 1742
Db 1758 ACTCCTTATGACCCTGAGGACCCAGGGCAAGAAACTAATGTGTCTATGTCTTTCATTTGG 1817
Qy 1743 cagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggctgag 1802
Db 1818 CAGTCTGCCCCAGACATTGGGAGAAAGTTAGAGAGGTTAGAAGATTTAAAAAACAAGACG 1877
Qy 1803 ttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagaggag 1862
Db 1878 CTTGGAGATTTGGTTAGAGAGGCAGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAA 1937
Qy 1863 aaggaacagagaaaagaaaaggagagagaagaaaggggaggaaagacgtgatagacggcaa 1922
Db 1938 AGAGAGGAACGTATCAGGAGAGAAACAGAGGAAAAAGAAGAACGCCGTAGGACAGAGGAT 1997
Qy 1923 gagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggagaga 1982
Db 1998 GAGCAGAAAGAGAAAGAAAGAGATCGTAGGAGACATAGAGAGATGAGCA------AGCTA 2051
Qy 1983 gagagagattttaggaaaattaggtcaggccctagacagtcagggaacctgggcaatagg 2042
Db 2052 TTGGCCACTGTCGTTAGTGGACAGAAACAGGATAGACAGGGAGGAGAAC---GAAGGAGG 2108
Qy 2043 accccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaaggaac 2102
Db 2109 TCCCAACTCGATCGCGACCAGTGTGCCTACTGCAAAGAAAAGGGGCACTGGGCTAAAGAT 2168
Qy 2103 tgccccaagaa-----gggaaacaaaggaccgaaggtcctagctctagaagaaga----t 2153
Db 2169 TGTCCCAAGAAACCACGAGGACCTCGGGGACCAAGACCCCAGACCTCCCTCCTGACCCTA 2228
Qy 2154 aaagattaggggagacggggttcggacccccctccccgagcccagggtaactttgaaggtg 2213
Db 2229 GATGACTAGGGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTC 2288
Qy 2214 gaggggcaaccagttgagttcctggttgataccggagcggagcattcagtgctgctacaa 2273
Db 2289 GGGGGGCAACCCGTCACCTTCCTGGTAGATACTGGGGCCCAACACTCCGTGCTGACCCAA 2348
Qy 2274 ccattaggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggcag 2333
Db 2349 AATCCTGGACCCCTAAGTGATAAGTCTGCCTGGGTCCAAGGGGCTACTGGAGGAAAGCGG 2408
Qy 2334 tatccatggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgttt 2393
Db 2409 TATCGCTGGACCACGGATCGCAAAGTACATCTAGCTACCGGTAAGGTCACCCACTCTTTC 2468
Qy 2394 ctggtcatccctgagtgcccagtacccttctaggtagagacttactgaccaagatggga 2453
Db 2469 CTCCATGTACCAGACTGTCCCTATCCTCTGTTAGGAAGAGATTTGCTGACTAAACTAAAA 2528
Qy 2454 gctcaaatttcttttgaacaaggaagaccagaagtgtctgtgaataaca---aacccatc 2510
Db 2529 GCCCAAATCCACTTTGAGGGATCAGGAGCTCAGGTTATGGGACCAATGGGGCAGCCCCTG 2588
Qy 2511 actgtgttgaccctccaattagatgatgaatatcgactatattctccccaagtaaagcct 2570
Db 2589 CAAGTGTTGACCCTAAATATAGAAGATGAGCATCGGCTACATGAGACCTCAAAAGAGCCA 2648
Qy 2571 gatcaagatatacagtcc---tggttggagcagtttccccaagcctgggcagaaaccgca 2627
Db 2649 GATGTTTCTCTAGGGTCCACATGGCTGTCTGATTTTCCTCAGGCCTGGGCGGAAACCGGG 2708
Qy 2628 gggatgggtttggcaaagcaagttcccccacaggttattcaactgaaggccagtgctaca 2687
Db 2709 GGCATGGGACTGGCAGTTCGCCAAGCTCCTCTGATCATACCTCTGAAAGCAACCTCTACC 2768
Qy 2688 ccagtatcagtcagacagtacccctgagtagagaggctcgagaaggaatttggccgcat 2747
Db 2769 CCCGTGTCCATAAAACAATACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCAC 2828
Qy 2748 gttcaaagattaatccaacagggcatcctagttcctgtccaatccccttggaatactccc 2807
Db 2829 ATACAGAGACTGTTGGACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCCC 2888



```
Qy    7064 ctgccctgatcacaggaccacagcagctagagaaaggacttggtgagctacatgcggcca 7123
           |   ||  || | | |  |       || |   | |    |||| || || ||
Db    7254 GAACAGGGACTACTGCTCTAATGGCCACTCAGCA---ATTCCAGCAGCTCCAAGCCGCAG 7310

Qy    7124 tgacagaagatctccgagccttaaaggagtctgttagcaacctagaagagtccctgactt 7183
           |    || |||||| | |   |  | | ||  |    ||||||||| |||| || ||||
Db    7311 TACAGGATGATCTCAGGGAGGTTGAAAAATCAATCTCTAACCTAGAAAAGTCTCTCACTT 7370

Qy    7184 ctttgtctgaagtggttctacagaaccggaggggattagatctgctgtttctaagagaag 7243
           | ||||||||||| || |||||||| || |||||  ||||  || | |||||||| |||||
Db    7371 CCCTGTCTGAAGTTGTCCTACAGAATCGAAGGGGCCTAGACTTGTTATTTCTAAAAGAAG 7430

Qy    7244 gtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcactcaggagcca 7303
           | ||| | ||||| ||  ||||||||||||||||||||||||  || ||| |||||
Db    7431 GAGGGCTGTGTGCTGCTCTAAAAGAAGAATGTTGCTTCTATGCGGACCACACAGGACTAG 7490

Qy    7304 tcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagagg 7363
           | ||||||  ||||  |||  | ||| | | | |  | | | |          |||
Db    7491 TGAGAGACAGCATGGCCAAATTGAGAGAGAGGCTTAATCAGAGACAGAAACTGTTTGAGT 7550

Qy    7364 ctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgaccaccctgcttt 7423
           |    || || |||||||| |||  ||| ||||| || |||||| | |||||| || | |
Db    7551 CAACTCAAGGATGGTTTGAGGGACTGTTTAACAGATCCCCTTGGTTTACCACCTTGATAT 7610

Qy    7424 ctgctctgacggggcccctagtagtcctgctcctgttacttacagttgggccttgcttaa 7483
           || |  | | ||| |||||  | || || ||  || |  |     | || || ||| |
Db    7611 CTACCATTATGGGACCCCTCATTGTACTCCTAATGATTTTGCTCTTCGGACCCTGCATTC 7670

Qy    7484 ttaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatcatggtactta 7543
           |||||  | || ||    ||||||| |||  |  |     | ||||||    | ||  | |
Db    7671 TTAATCGATTAGTCCAATTTGTTAAAGACAGGATATCAGTGGTCCAGGCTCTAGTTTTGA 7730

Qy    7544 ggcaacagtacca 7556
             |||||  || ||
Db    7731 CTCAACAATATCA 7743
```

```
RESULT  12
US-08-110-300A-9/c
; Sequence 9, Application US/08110300A
; Patent No. 5643756
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,300A
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-110-300A-9
```

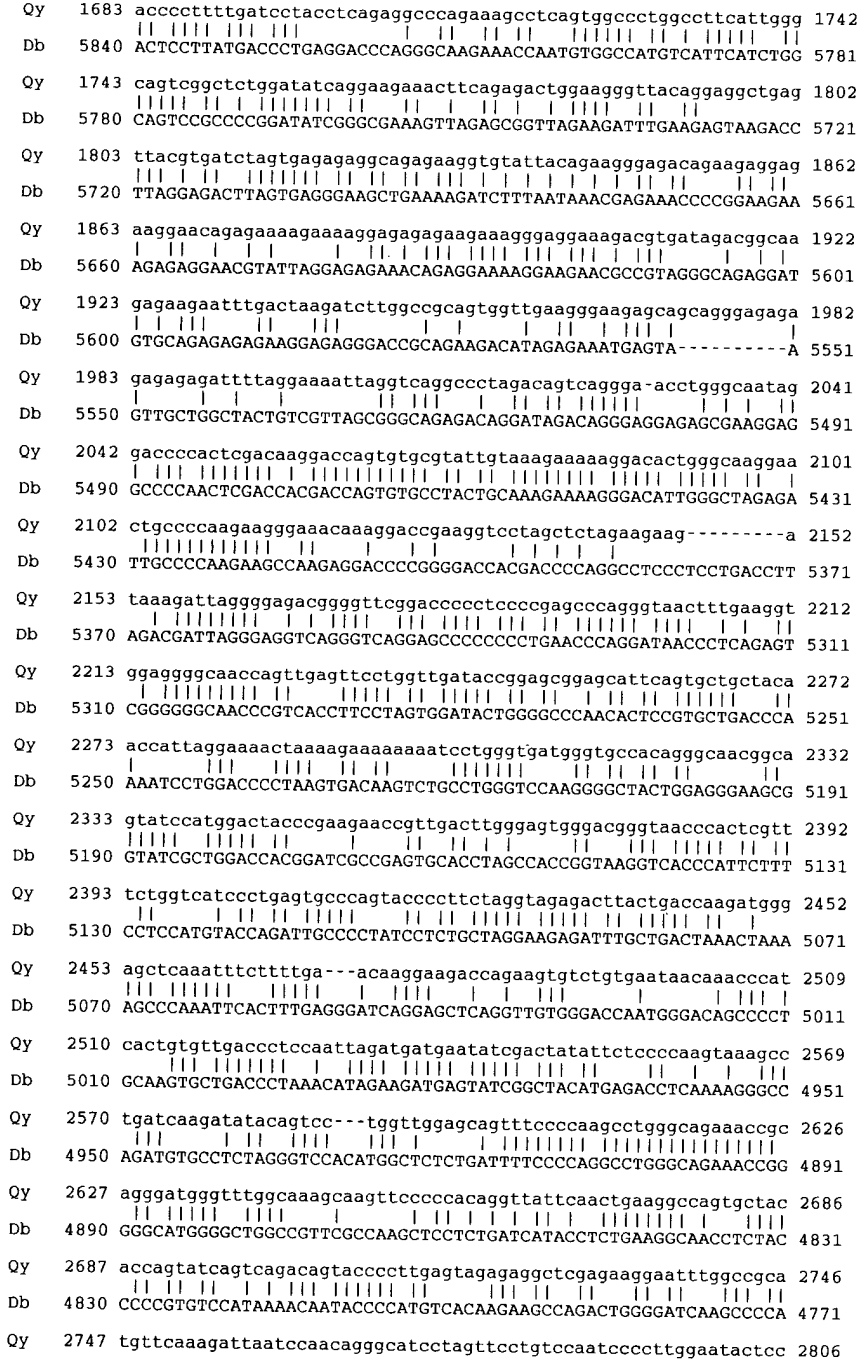
```
  Query Match             19.4%;  Score 1596;  DB 1;  Length 10367;
  Best Local Similarity   58.4%;  Pred. No. 0;
  Matches 3161;  Conservative    0;  Mismatches 2135;  Indels  114;  Gaps   17;

Qy     467 tccgcggccgtccgactcttttgcctgcttgtggaa--gacgcggacgggtcgcgtgtgt 524
           |||||  ||||| || | |||    |   | |||||  || || |    |       |||
Db    7098 TCCGCCCCCGTCTGAGTTTTTGCTTTCGGTTTGGAACCGAAGCCGCGCCGCGCGTCTTGT 7039

Qy     525 ctggatctgttggtttctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagtttt 584
           |||   | |     |||||| | || | | |||| ||  |   | ||    |   |
Db    7038 CTGCTGCAGCATCGTTCTGTGTTGTCTCTGTTTGACTGTT--TTTCTGTATTTGTCTGAA 6981

Qy     585 aatatgggacagacagtgactacccccctagtttgactctcgaccattggactgaagtt 644
           || ||||| ||| | || || |||||| | |||||||| | ||||| ||||  || ||
Db    6980 AACATGGGCCAGGCTGTTACCACCCCCTTAAGTTTGACTTTAGACCACTGGAAGGATGTC 6921

Qy     645 agatccagggctcataatttgtcagttcaggttaagaagggaccttggcagactttctgt 704
             |   |  || || ||  |||| ||  ||||||  ||  | |  |||   || |||||
Db    6920 GAACGGACAGCCCACAACCTGTCGGTAGAGGTTAGAAAAAGGCGCTGGGTTACATTCTGC 6861

Qy     705 gcctctgaatggccaacattcgatgttggatggccatcagaggggacctttaattctgaa 764
           |  | ||||||||||| ||| | || |||||||||  ||| || || ||||| | ||
Db    6860 TCTGCAGAATGGCCAACCTTCAACGTCGGATGGCCACGAGACGGCACTTTTAACCCAGAC 6801

Qy     765 attatcctggctgttaaggcaatcatttttcagactggacccggctctcatcctgatcag 824
           |||||       ||||||  |   | ||     |||| ||       ||||| ||||||
Db    6800 ATTATTACACAGGTTAAGATCAAGGTCTTCTCACCTGGCCCACATGGACATCCGGATCAG 6741

Qy     825 gagccctatatccttacgtggcaagatttggcagaagatcctccgccatgggttaaacca 884
           |  ||||| ||| | || ||| ||| | | |||| ||| || || || ||||| | |||
Db    6740 GTCCCCTACATCGTGACCTGGGAAGCTATAGCAGTAGACCCCCCTCCCTGGGTCAGACCC 6681

Qy     885 tggctaaataaa-------------------ccaagaaagccaggtccccgaatcctggc 925
           |   |  |                     ||     ||||  || |    ||||
Db    6680 TTCGTGCACCCTAAACCTCCCCTCTCTCTTCCCCCTTCAGCCCCCTCTCTCCCACCTGAA 6621

Qy     926 tcttggagagaaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctacc 985
           |           |  |  | |  |          ||  || |||          |  | |
Db    6620 CCCCCACTCTCGACCCCGCCCCAGTCCTCCCTCTATCCGGCTCTCACTTCTCCTTTAAAC 6561

Qy     986 ccgagatcgaggagccgccgacttggccggaaccccaacctgttcccccaccccccttatc 1045
           | | | | |||   |     |||      | |      |   ||    | |  || |
Db    6560 ACCAAACCTAGGCCTCAAGTCCTTCCTGATAGCGGAGGACCACTCATTGATCTACTCACG 6501

Qy    1046 cagcacagggtgctgtgaggggacctctgcccctcctggagctccggtggtggagggacc 1105
           || ||    ||  |  ||||||   | ||  |||       |||    || | | |
Db    6500 GAGGACCCTCCGCCTTACCGGGACCCAGGGCCACCCTCTCCTGACGGGAACGGCGATAGC 6441

Qy    1106 -------tgctgccgggactcggagccggagaggcgccacccccggagcggacagacgaga 1158
                  ||   ||   |  ||||||   ||  | | ||   |     |   | |
Db    6440 GGAGAAGTGGCCCCTACAGAAGGAGCCCCTGACCCTTCCCCAATGGTATCCCGCCTGCGG 6381

Qy    1159 tcgcgatattaccgctgcgcacctatggccctcccatgccaggggccaattgcagccc- 1217
             || |  | | |  ||   |      |     | |  |||| |  |       |||
Db    6380 GGAAGAAAAGAACCCCCCGTGGCGGATTCTACTACCTCTCAGGCGTTCCCCCTTCGCCTG 6321

Qy    1218 --------------ctccagtattggccctttttcttctgcagatctctataattggaaa 1262
                           || || ||||| ||||| ||  | || ||||||| ||||||
Db    6320 GGAGGGAATGGACAGTATCAATACTGGCCATTTTCCTCCTCTGACCTCTATAACTGGAAA 6261

Qy    1263 actaaccatcccccttttctcggaggatccccaacgcctcacggggttggtggagtccctt 1322
           | |||| | ||| ||||||| ||||| ||        | || |  ||| | |||||| ||
Db    6260 AATAACAACCCCCTCTTTCTCCGAGGACCCAGCTAAATTGACAGCTTTGATCGAGTCCGTT 6201

Qy    1323 atgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttcaca 1382
```

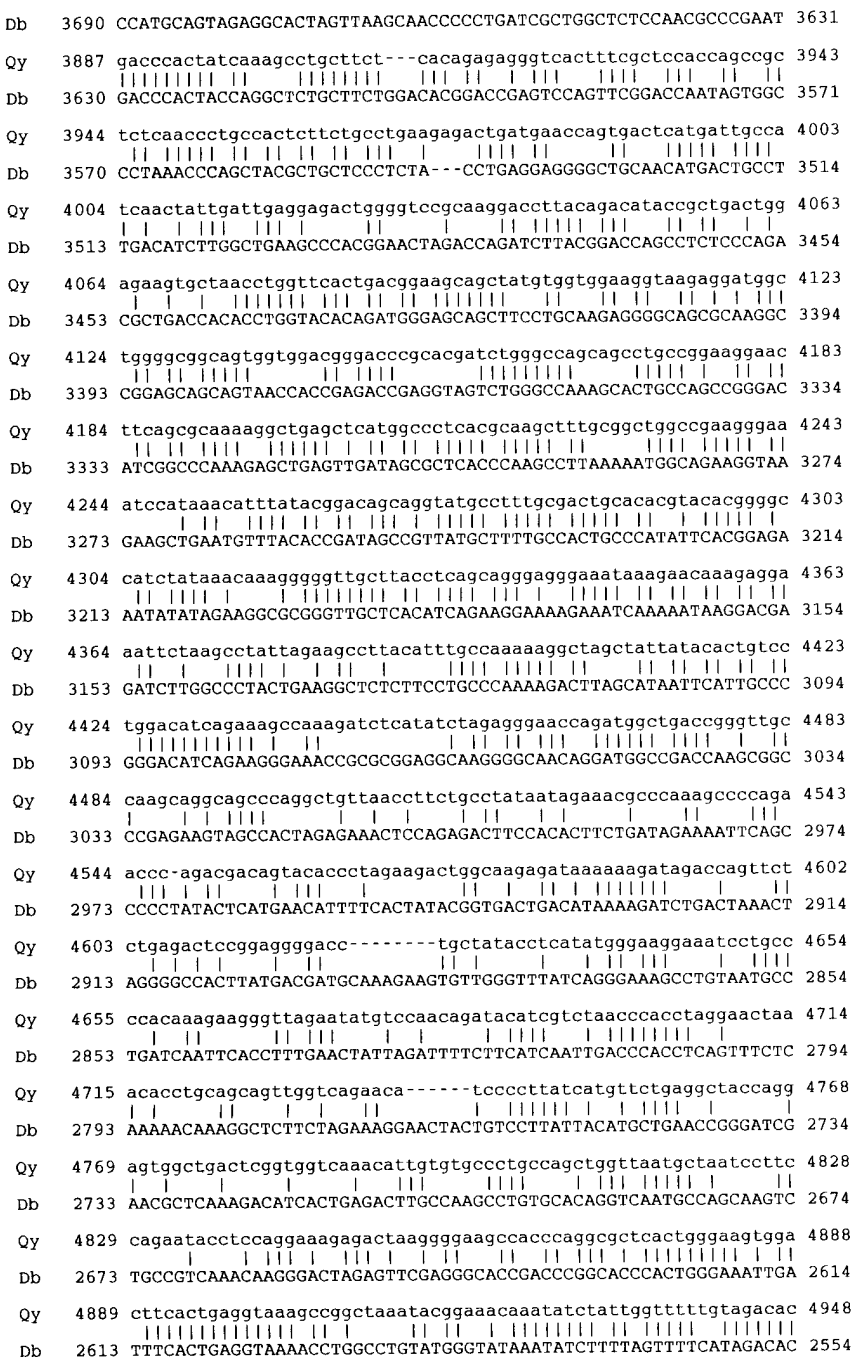

Db 6200 CTCCTTACTCATCAGCCCACTTGGGATGACTGCCAACAGCTATTAGGGACCCTGCTGACG 6141
Qy 1383 accgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgacggg 1442
Db 6140 GGAGAAGAAAAACAGCGAGTGCTCCTAGAGGCCCGAAAGGCGGTTCGAGGGGAGGACGGA 6081
Qy 1443 cgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggttgg 1502
Db 6080 CGCCCAACTCAGCTGCCCAATGACATTAATGATGCTTTTCCCTTGGAACGTCCCGACTGG 6021
Qy 1503 gactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcg 1562
Db 6020 GACTACAACACCCAACGAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCG 5961
Qy 1563 ggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcag 1622
Db 5960 GGTCTCCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGGATAACCCAG 5901
Qy 1623 ggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcggttc 1682
Db 5900 GGACCTAATGAGTCTCCCTCAGCCTTTTTAGAGAGACTCAAGGAGGCCTATCGCAGATAC 5841
Qy 1683 acccctttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcattggg 1742
Db 5840 ACTCCTTATGACCCTGAGGACCCAGGGCAAGAAACCAATGTGGCCATGTCATTCATCTGG 5781
Qy 1743 cagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggctgag 1802
Db 5780 CAGTCCGCCCCGGATATCGGGCGAAAGTTAGAGCGGTTAGAAGATTTGAAGAGTAAGACC 5721
Qy 1803 ttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagaggag 1862
Db 5720 TTAGGAGACTTAGTGAGGGAAGCTGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAA 5661
Qy 1863 aaggaacagagaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacggcaa 1922
Db 5660 AGAGAGGAACGTATTAGGAGAGAAACAGAGGAAAAGGAAGAACGCCGTAGGGCAGAGGAT 5601
Qy 1923 gagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggagaga 1982
Db 5600 GTGCAGAGAGAGAAGGAGAGGGACCGCAGAAGACATAGAGAAATGAGTA----------A 5551
Qy 1983 gagagagattttaggaaaattaggtcaggccctagacagtcaggga-acctgggcaatag 2041
Db 5550 GTTGCTGGCTACTGTCGTTAGCGGGCAGAGACAGGATAGACAGGGAGGAGAGCGAAGGAG 5491
Qy 2042 gaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaaggaa 2101
Db 5490 GCCCCAACTCGACCACGACCAGTGTGCCTACTGCAAAGAAAAGGGACATTGGGCTAGAGA 5431
Qy 2102 ctgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaag---------a 2152
Db 5430 TTGCCCCAAGAAGCCAAGAGGACCCCGGGGACCACGACCCCAGGCCTCCCTCCTGACCTT 5371
Qy 2153 taaagattaggggagacggggttcggacccccctccccgagcccagggtaactttgaaggt 2212
Db 5370 AGACGATTAGGGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAGAGT 5311
Qy 2213 ggaggggcaaccagttgagttcctggttgataccggagcggagcattcagtgctgctaca 2272
Db 5310 CGGGGGGCAACCCGTCACCTTCCTAGTGGATACTGGGGCCCAACACTCCGTGCTGACCCA 5251
Qy 2273 accattaggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggca 2332
Db 5250 AAATCCTGGACCCCTAAGTGACAAGTCTGCCTGGGTCCAAGGGGCTACTGGAGGGAAGCG 5191
Qy 2333 gtatccatggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtt 2392
Db 5190 GTATCGCTGGACCACGGATCGCCGAGTGCACCTAGCCACCGGTAAGGTCACCCATTCTTT 5131
Qy 2393 tctggtcatccctgagtgcccagtacccttctaggtagagacttactgaccaagatggg 2452
Db 5130 CCTCCATGTACCAGATTGCCCCTATCCTCTGCTAGGAAGAGATTTGCTGACTAAACTAAA 5071
Qy 2453 agctcaaatttcttttga---acaaggaagaccagaagtgtctgtgaataacaaacccat 2509
Db 5070 AGCCCAAATTCACTTTGAGGGATCAGGAGCTCAGGTTGTGGGACCAATGGGACAGCCCCT 5011
Qy 2510 cactgtgttgaccctccaattagatgatgaatatcgactatattctccccaagtaaagcc 2569
Db 5010 GCAAGTGCTGACCCTAAACATAGAAGATGAGTATCGGCTACATGAGACCTCAAAAGGGCC 4951
Qy 2570 tgatcaagatatacagtcc---tggttggagcagtttccccaagcctgggcagaaaccgc 2626
Db 4950 AGATGTGCCTCTAGGGTCCACATGGCTCTCTGATTTTCCCCAGGCCTGGGCAGAAACCGG 4891
Qy 2627 agggatgggtttggcaaagcaagttcccccacaggttattcaactgaaggccagtgctac 2686
Db 4890 GGGCATGGGGCTGGCCGTTCGCCAAGCTCCTCTGATCATACCTCTGAAGGCAACCTCTAC 4831
Qy 2687 accagtatcagtcagacagtacccccttgagtagagaggctcgagaaggaatttggccgca 2746
Db 4830 CCCCGTGTCCATAAAACAATACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCA 4771
Qy 2747 tgttcaaagattaatccaacagggcatcctagttcctgtccaatccccttggaatactcc 2806
Db 4770 CATACAGAGACTGCTGGATCAGGGAATTCTGGTACCCTGCCAGTCCCCCTGGAACACGCC 4711
Qy 2807 cctgctaccggttaggaagcctgggaccaatgattatcgaccagtacaggacttgagaga 2866
Db 4710 CCTGCTACCCGTTAAGAAACCGGGGACTAATGATTATAGGCCTGTCCAGGATCTGAGAGA 4651
Qy 2867 ggtcaataaaagggtgcaggacatacacccaacggtcccgaaccctataacctcttgag 2926
Db 4650 AGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCCAACCCTTACAACCTCTTGAG 4591
Qy 2927 cgccctcccgcctgaacggaactggtacacagtattggacttaaaagatgccttcttctg 2986
Db 4590 CGGGCTCCCACCGTCCCACCAGTGGTACACTGTGCTTGACTTAAAAGATGCTTTTTTCTG 4531
Qy 2987 cctgagattacaccccactagccaaccacttttgccttcgaatggagagatccaggtac 3046
Db 4530 CCTGAGACTCCACCCCACCAGTCAGTCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGAT 4471
Qy 3047 gggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactccccgac 3106
Db 4470 GGGAATCTCAGGACAATTAACCTGGACCAGACTCCCGCAGGGTTTCAAAAACAGTCCCAC 4411
Qy 3107 catctttgacgaagccctacacagggacctggccaacttcaggatccaacaccctcaggt 3166
Db 4410 CCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGGATCCAGCACCCAGACCT 4351
Qy 3167 gaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggactgctt 3226
Db 4350 GATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCCACTTCTGAGCTTGACTGTCA 4291
Qy 3227 agaaggtacgaaggcactactgctggaattgtctgacctaggctacagagcctctgctaa 3286
Db 4290 ACAAGGTACGCGGGCCCTGTTACAAACCCTAGGGGACCTCGGATATCGGGCCTCGGCCAA 4231
Qy 3287 gaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggcgggca 3346
Db 4230 GAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCA 4171
Qy 3347 gcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaaccacagc 3406
Db 4170 GAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 4111
Qy 3407 caaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccggggtt 3466
Db 4110 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGATCCCTGGGTT 4051
Qy 3467 tgcgaccttagcagccccactctacccgctaaccaaagaaaaaggggattctcctgggc 3526
Db 4050 TGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACGGGGACTCTGTTTGAGTGGGG 3991

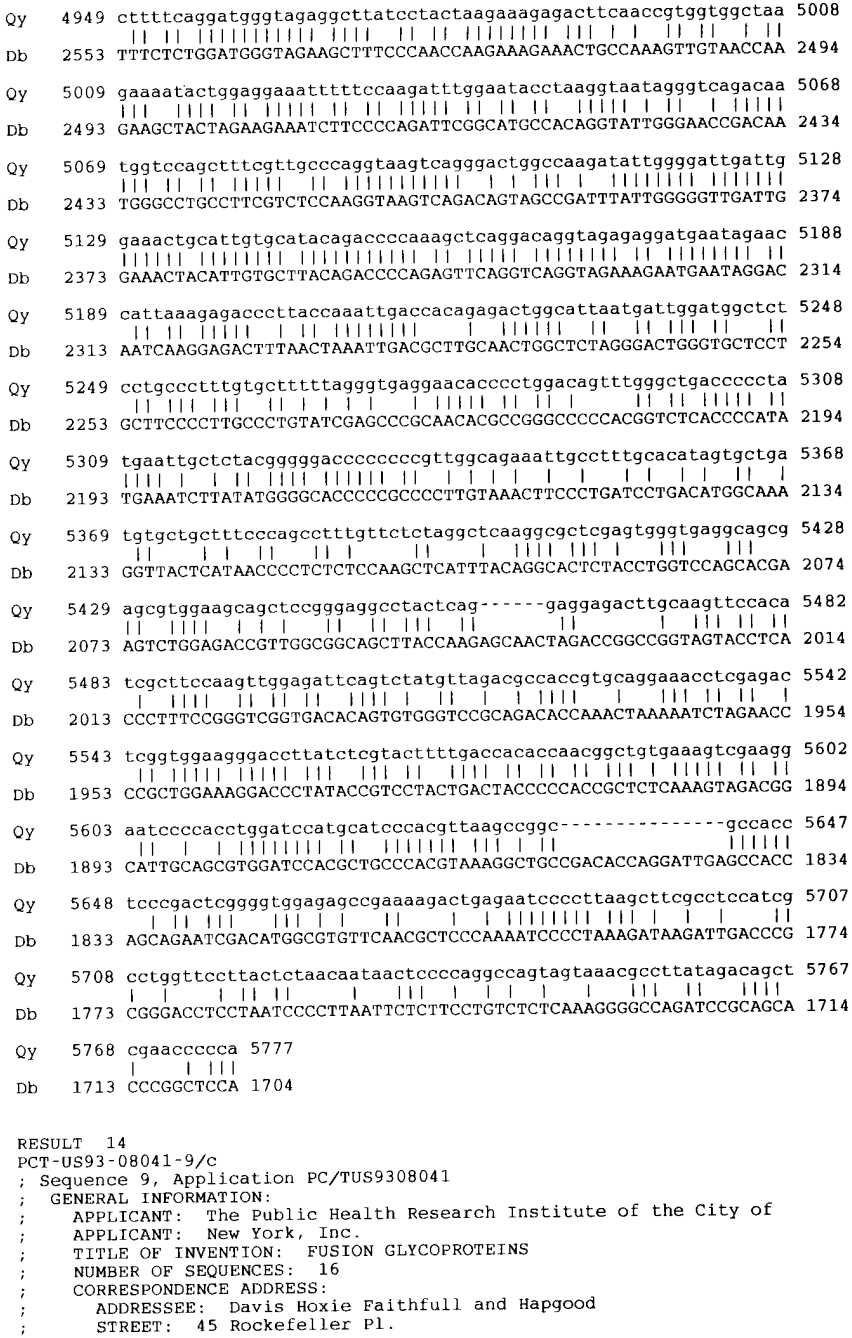
```
Qy   3527 tcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctct 3586
Db   3990 CCCAGACCAGCAAAAGGCCTACCAAGAGATCAAGCAGGCTCTCTTAACTGCCCCTGCCCT 3931

Qy   3587 ggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaagggagtagc 3646
Db   3930 GGGATTGCCAGACTTGACTAAGCCCTTCGAACTTTTTGTTGACGAGAAGCAGGGCTACGC 3871

Qy   3647 ccgaggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctgtcaaa 3706
Db   3870 CAAAGGTGTCCTAACGCAAAAACTGGGGCCTTGGCGTCGGCCGGTGGCCTACCTGTCCAA 3811

Qy   3707 gaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggc 3766
Db   3810 AAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAGCAGCCATCGC 3751

Qy   3767 catactggtcaaggacgctgacaaattgactttgggacagaatataactgtaatagcccc 3826
Db   3750 CGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAGCCACTAGTCATTCTGGCCCC 3691

Qy   3827 ccatgcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgcccgcat 3886
Db   3690 CCATGCAGTAGAGGCACTAGTTAAGCAACCCCCTGATCGCTGGCTCTCCAACGCCCGAAT 3631

Qy   3887 gacccactatcaaagcctgcttct---cacagagagggtcactttcgctccaccagccgc 3943
Db   3630 GACCCACTACCAGGCTCTGCTTCTGGACACGGACCGAGTCCAGTTCGGACCAATAGTGGC 3571

Qy   3944 tctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgcca 4003
Db   3570 CCTAAACCCAGCTACGCTGCTCCCTCTA---CCTGAGGAGGGGCTGCAACATGACTGCCT 3514

Qy   4004 tcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctgactgg 4063
Db   3513 TGACATCTTGGCTGAAGCCCACGGAACTAGACCAGATCTTACGGACCAGCCTCTCCCAGA 3454

Qy   4064 agaagtgctaacctggttcactgacggaagcagctatgtggtggaaggtaagaggatggc 4123
Db   3453 CGCTGACCACACCTGGTACACAGATGGGAGCAGCTTCCTGCAAGAGGGGCAGCGCAAGGC 3394

Qy   4124 tggggcggcagtggtggacgggacccgcacgatctgggccagcagcctgccggaaggaac 4183
Db   3393 CGGAGCAGCAGTAACCACCGAGACCGAGGTAGTCTGGGCCAAAGCACTGCCAGCCGGGAC 3334

Qy   4184 ttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaagggaa 4243
Db   3333 ATCGGCCCAAAGAGCTGAGTTGATAGCGCTCACCCAAGCCTTAAAAATGGCAGAAGGTAA 3274

Qy   4244 atccataaacatttatacggacagcaggtatgcctttgcgactgcacacgtacacggggc 4303
Db   3273 GAAGCTGAATGTTTACACCGATAGCCGTTATGCTTTTGCCACTGCCCATATTCACGGAGA 3214

Qy   4304 catctataaacaaaggggttgcttacctcagcagggagggaaataaagaacaaagagga 4363
Db   3213 AATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAAATCAAAAATAAGGACGA 3154

Qy   4364 aattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgtcc 4423
Db   3153 GATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGACTTAGCATAATTCATTGCCC 3094

Qy   4424 tggacatcagaaagccaaagatctcatatctagagggaaccagatggctgaccgggttgc 4483
Db   3093 GGGACATCAGAAGGGAAACCGCGCGGAGGCAAGGGGCAACAGGATGGCCGACCAAGCGGC 3034

Qy   4484 caagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaaagccccaga 4543
Db   3033 CCGAGAAGTAGCCACTAGAGAAACTCCAGAGACTTCCACACTTCTGATAGAAAATTCAGC 2974

Qy   4544 accc-agacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttct 4602
Db   2973 CCCCTATACTCATGAACATTTTCACTATACGGTGACTGACATAAAAGATCTGACTAAACT 2914
```

```
Qy   4603 ctgagactccggaggggacc--------tgctatacctcatatgggaaggaaatcctgcc 4654
Db   2913 AGGGGCCACTTATGACGATGCAAAGAAGTGTTGGGTTTATCAGGGAAAGCCTGTAATGCC 2854

Qy   4655 ccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaa 4714
Db   2853 TGATCAATTCACCTTTGAACTATTAGATTTTCTTCATCAATTGACCCACCTCAGTTTCTC 2794

Qy   4715 acacctgcagcagttggtcagaaca------tccccttatcatgttctgaggctaccagg 4768
Db   2793 AAAAACAAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTACATGCTGAACCGGGATCG 2734

Qy   4769 agtggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttc 4828
Db   2733 AACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAGGTCAATGCCAGCAAGTC 2674

Qy   4829 cagaatacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtgga 4888
Db   2673 TGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCGACCCGGCACCCACTGGGAAATTGA 2614

Qy   4889 cttcactgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacac 4948
Db   2613 TTTCACTGAGGTAAAACCTGGCCTGTATGGGTATAAATATCTTTTAGTTTTCATAGACAC 2554

Qy   4949 cttttcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaa 5008
Db   2553 TTTCTCTGGATGGGTAGAAGCTTTCCCAACCAAGAAAGAAACTGCCAAAGTTGTAACCAA 2494

Qy   5009 gaaaatactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaa 5068
Db   2493 GAAGCTACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAGGTATTGGGAACCGACAA 2434

Qy   5069 tggtccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattg 5128
Db   2433 TGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGATTTATTGGGGGTTGATTG 2374

Qy   5129 gaaactgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaac 5188
Db   2373 GAAACTACATTGTGCTTACAGACCCCAGAGTTCAGGTCAGGTAGAAAGAATGAATAGGAC 2314

Qy   5189 cattaaagagaccctaccaaattgaccacagagactggcattaatgattggatggctct 5248
Db   2313 AATCAAGGAGACTTTAACTAAATTGACGCTTGCAACTGGCTCTAGGGACTGGGTGCTCCT 2254

Qy   5249 cctgccctttgtgcttttagggtgaggaacacccctggacagtttgggctgacccccta 5308
Db   2253 GCTTCCCCTTGCCCTGTATCGAGCCCGCAACACGCCGGGCCCCCACGGTCTCACCCCATA 2194

Qy   5309 tgaattgctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctga 5368
Db   2193 TGAAATCTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGATCCTGACATGGCAAA 2134

Qy   5369 tgtgctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcg 5428
Db   2133 GGTTACTCATAACCCCTCTCTCCAAGCTCATTTACAGGCACTCTACCTGGTCCAGCACGA 2074

Qy   5429 agcgtggaagcagctccgggaggcctactcag------gaggagacttgcaagttccaca 5482
Db   2073 AGTCTGGAGACCGTTGGCGGCAGCTTACCAAGAGCAACTAGACCGGCCGGTAGTACCTCA 2014

Qy   5483 tcgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagac 5542
Db   2013 CCCTTTCCGGGTCGGTGACACAGTGTGGGTCCGCAGACACCAAACTAAAAATCTAGAACC 1954

Qy   5543 tcggtggaagggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaagg 5602
Db   1953 CCGCTGGAAAGGACCCTATACCGTCCTACTGACTACCCCCACCGCTCTCAAAGTAGACGG 1894

Qy   5603 aatccccacctggatccatgcatcccacgttaagccggc----------------gccacc 5647
Db   1893 CATTGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGACACCAGGATTGAGCCACC 1834

Qy   5648 tcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcgcctccatcg 5707
```

```
Db   1833 AGCAGAATCGACATGGCGTGTTCAACGCTCCCAAAATCCCCTAAAGATAAGATTGACCCG 1774

Qy   5708 cctggttccttactctaacaataactccccaggccagtagtaaacgccttatagacagct 5767

Db   1773 CGGGACCTCCTAATCCCCTTAATTCTCTTCCTGTCTCTCAAAGGGGCCAGATCCGCAGCA 1714

Qy   5768 cgaaccccca 5777

Db   1713 CCCGGCTCCA 1704


RESULT 13
US-08-886-642-9/c
; Sequence 9, Application US/08886642
; Patent No. 5952474
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  45 Rockefeller Plaza, Suite 2800
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/886,642
;      FILING DATE:  01-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/110,300
;      FILING DATE:  20-AUG-1993
;      APPLICATION NUMBER:  07/938,100
;      FILING DATE:  28-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hone, William J.
;      REGISTRATION NUMBER:  26,739
;      REFERENCE/DOCKET NUMBER:  07763/010002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212/765-5070
;      TELEFAX:  212/258-2291
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-886-642-9


  Query Match             19.4%;  Score 1596;  DB 2;  Length 10367;
  Best Local Similarity   58.4%;  Pred. No. 0;
  Matches 3161;  Conservative    0;  Mismatches 2135;  Indels  114;  Gaps   17;

Qy    467 tccgcggccgtccgactcttttgcctgcttgtggaa--gacgcggacgggtcgcgtgtgt 524

Db   7098 TCCGCCCCCGTCTGAGTTTTTGCTTTCGGTTTGGAACCGAAGCCGCGCCGCGCGTCTTGT 7039

Qy    525 ctggatctgttggtttctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagtttt 584

Db   7038 CTGCTGCAGCATCGTTCTGTGTTGTCTCTGTTTGACTGTT--TTTCTGTATTTGTCTGAA 6981

Qy    585 aatatgggacagacagtgactacccccttagtttgactctcgaccattggactgaagtt 644

Db   6980 AACATGGGCCAGGCTGTTACCACCCCCTTAAGTTTGACTTTAGACCACTGGAAGGATGTC 6921

Qy    645 agatccagggctcataatttgtcagttcaggttaagaagggaccttggcagactttctgt 704

Db   6920 GAACGGACAGCCCACAACCTGTCGGTAGAGGTTAGAAAAAGGCGCTGGGTTACATTCTGC 6861

Qy    705 gcctctgaatggccaacattcgatgttggatggccatcagaggggacctttaattctgaa 764

Db   6860 TCTGCAGAATGGCCAACCTTCAACGTCGGATGGCCACGAGACGGCACTTTTAACCCAGAC 6801

Qy    765 attatcctggctgttaaggcaatcatttttcagactggacccggctctcatcctgatcag 824

Db   6800 ATTATTACACAGGTTAAGATCAAGGTCTTCTCACCTGGCCCACATGGACATCCGGATCAG 6741

Qy    825 gagccctatatccttacgtggcaagatttggcagaagatcctccgccatgggttaaacca 884

Db   6740 GTCCCCTACATCGTGACCTGGGAAGCTATAGCAGTAGACCCCCCTCCCTGGGTCAGACCC 6681

Qy    885 tggctaaataaa----------ccaagaaagccaggtccccgaatcctggc 925

Db   6680 TTCGTGCACCCTAAACCTCCCCTCTCTCTTCCCCCTTCAGCCCCCTCTCTCCCACCTGAA 6621

Qy    926 tcttggagagaaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctacc 985

Db   6620 CCCCCACTCTCGACCCCGCCCCAGTCCTCCCTCTATCCGGCTCTCACTTCTCCTTTAAAC 6561

Qy    986 ccgagatcgaggagccgccgacttggccggaaccccaacctgttcccccaccccttatc 1045

Db   6560 ACCAAACCTAGGCCTCAAGTCCTTCCTGATAGCGGAGGACCACTCATTGATCTACTCACG 6501

Qy   1046 cagcacagggtgctgtgagggacctctgcccctcctggagctccggtggtggagggacc 1105

Db   6500 GAGGACCCTCCGCCTTACCGGGACCCAGGGCCACCCTCTCCTGACGGGAACGGCGATAGC 6441

Qy   1106 -------tgctgccgggactcggagccggagaggcgccaccccggagcggacagacgaga 1158

Db   6440 GGAGAAGTGGCCCCTACAGAAGGAGCCCCTGACCCTTCCCCAATGGTATCCCGCCTGCGG 6381

Qy   1159 tcgcgatattaccgctgcgcacctatggccctcccatgccaggggggccaattgcagccc- 1217

Db   6380 GGAAGAAAAGAACCCCCCGTGGCGGATTCTACTACCTCTCAGGCGTTCCCCCTTCGCCTG 6321

Qy   1218 ---------------ctccagtattggccctttttcttctgcagatctctataattggaaa 1262

Db   6320 GGAGGGAATGGACAGTATCAATACTGGCCATTTTCCTCCTCTGACCTCTATAACTGGAAA 6261

Qy   1263 actaaccatcccccttttctcggaggatccccaacgcctcacggggttggtggagtccctt 1322

Db   6260 AATAACAACCCCTCTTTCTCCGAGGACCCAGCTAAATTGACAGCTTTGATCGAGTCCGTT 6201

Qy   1323 atgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttcaca 1382

Db   6200 CTCCTTACTCATCAGCCCACTTGGGATGACTGCCAACAGCTATTAGGGACCCTGCTGACG 6141

Qy   1383 accgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgacggg 1442

Db   6140 GGAGAAGAAAAACAGCGAGTGCTCCTAGAGGCCCGAAAGGCGGTTCGAGGGGAGGACGGA 6081

Qy   1443 cgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggttgg 1502

Db   6080 CGCCCAACTCAGCTGCCCAATGACATTAATGATGCTTTTCCCTTGGAACGTCCCGACTGG 6021

Qy   1503 gactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcg 1562

Db   6020 GACTACAACACCCAACGAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCG 5961

Qy   1563 ggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcag 1622

Db   5960 GGTCTCCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGGATAACCCAG 5901

Qy   1623 ggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcggttc 1682

Db   5900 GGACCTAATGAGTCTCCCTCAGCCTTTTTAGAGAGACTCAAGGAGGCCTATCGCAGATAC 5841
```


Qy 1683 acccctttttgatcctacctcagaggcccagaaagcctcagtggccctggccttcattggg 1742
Db 5840 ACTCCTTATGACCCTGAGGACCCAGGGCAAGAAACCAATGTGGCCATGTCATTCATCTGG 5781
Qy 1743 cagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggctgag 1802
Db 5780 CAGTCCGCCCCGGATATCGGGCGAAAGTTAGAGCGGTTAGAAGATTTGAAGAGTAAGACC 5721
Qy 1803 ttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagaggag 1862
Db 5720 TTAGGAGACTTAGTGAGGGAAGCTGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAA 5661
Qy 1863 aaggaacagagaaaagaaaaggagagagaagaaagggaggaaaagacgtgatagacggcaa 1922
Db 5660 AGAGAGGAACGTATTAGGAGAGAAACAGAGGAAAAGGAAGAACGCCGTAGGGCAGAGGAT 5601
Qy 1923 gagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggagaga 1982
Db 5600 GTGCAGAGAGAGAAGGAGAGGGACCGCAGAAGACATAGAGAAATGAGTA----------A 5551
Qy 1983 gagagagattttaggaaaattaggtcaggccctagacagtcaggga-acctgggcaatag 2041
Db 5550 GTTGCTGGCTACTGTCGTTAGCGGGCAGAGACAGGATAGACAGGGAGGAGAGCGAAGGAG 5491
Qy 2042 gaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggcaaggaa 2101
Db 5490 GCCCCAACTCGACCACGACCAGTGTGCCTACTGCAAAGAAAAGGGACATTGGGCTAGAGA 5431
Qy 2102 ctgccccaagaagggaaacaaaggaccgaaggtcctagctctagaagaag---------a 2152
Db 5430 TTGCCCCAAGAAGCCAAGAGGACCCCGGGGACCACGACCCCAGGCCTCCCTCCTGACCTT 5371
Qy 2153 taaagattaggggagacggggttcggaccccctccccgagcccagggtaactttgaaggt 2212
Db 5370 AGACGATTAGGGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAGAGT 5311
Qy 2213 ggaggggcaaccagttgagttcctggttgataccggagcggagcattcagtgctgctaca 2272
Db 5310 CGGGGGGCAACCCGTCACCTTCCTAGTGGATACTGGGGCCCAACACTCCGTGCTGACCCA 5251
Qy 2273 accattaggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggcaacggca 2332
Db 5250 AAATCCTGGACCCCTAAGTGACAAGTCTGCCTGGGTCCAAGGGGCTACTGGAGGGAAGCG 5191
Qy 2333 gtatccatggactacccgaagaaccgttgacttgggagtgggacgggtaacccactcgtt 2392
Db 5190 GTATCGCTGGACCACGGATCGCCGAGTGCACCTAGCCACCGGTAAGGTCACCCATTCTTT 5131
Qy 2393 tctggtcatccctgagtgcccagtacccttctaggtagagacttactgaccaagatggg 2452
Db 5130 CCTCCATGTACCAGATTGCCCCTATCCTCTGCTAGGAAGAGATTTGCTGACTAAACTAAA 5071
Qy 2453 agctcaaatttcttttga---acaaggaagaccagaagtgtctgtgaataacaaacccat 2509
Db 5070 AGCCCAAATTCACTTTGAGGGATCAGGAGCTCAGGTTGTGGGACCAATGGGACAGCCCCT 5011
Qy 2510 cactgtgttgaccctccaattagatgatgaatatcgactatattctccccaagtaaagcc 2569
Db 5010 GCAAGTGCTGACCCTAAACATAGAAGATGAGTATCGGCTACATGAGACCTCAAAAGGGCC 4951
Qy 2570 tgatcaagatatacagtcc---tggttggagcagtttccccaagcctgggcagaaaccgc 2626
Db 4950 AGATGTGCCTCTAGGGTCCACATGGCTCTCTGATTTTCCCCAGGCCTGGGCAGAAACCGG 4891
Qy 2627 agggatgggtttggcaaagcaagttcccccacaggttattcaactgaaggccagtgctac 2686
Db 4890 GGGCATGGGGCTGGCCGTTCGCCAAGCTCCTCTGATCATACCTCTGAAGGCAACCTCTAC 4831
Qy 2687 accagtatcagtcagacagtacccctgagtagagaggctcgagaaggaatttggccgca 2746
Db 4830 CCCCGTGTCCATAAAACAATACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCA 4771
Qy 2747 tgttcaaagattaatccaacagggcatcctagttcctgtccaatccccttggaatactcc 2806
Db 4770 CATACAGAGACTGCTGGATCAGGGAATTCTGGTACCCTGCCAGTCCCCCTGGAACACGCC 4711
Qy 2807 cctgctaccggttaggaagcctgggaccaatgattatcgaccagtacaggacttgagaga 2866
Db 4710 CCTGCTACCCGTTAAGAAACCGGGGACTAATGATTATAGGCCTGTCCAGGATCTGAGAGA 4651
Qy 2867 ggtcaataaaagggtgcaggacatacacccaacggtcccgaaccccttataacctcttgag 2926
Db 4650 AGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCCAACCCTTACAACCTCTTGAG 4591
Qy 2927 cgccctcccgcctgaacggaactggtacacagtattggacttaaaagatgccttcttctg 2986
Db 4590 CGGGCTCCCACCGTCCCACCAGTGGTACACTGTGCTTGACTTAAAAGATGCTTTTTTCTG 4531
Qy 2987 cctgagattacaccccactagccaaccacttttgccttcgaatggagagatccaggtac 3046
Db 4530 CCTGAGACTCCACCCCACCAGTCAGTCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGAT 4471
Qy 3047 gggaagaaccgggcagctcacctggacccgactgccccaagggttcaagaactccccgac 3106
Db 4470 GGGAATCTCAGGACAATTAACCTGGACCAGACTCCCGCAGGGTTTCAAAAACAGTCCCAC 4411
Qy 3107 catctttgacgaagccctacacagggacctggccaacttcaggatccaacaccctcaggt 3166
Db 4410 CCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGGATCCAGCACCCAGACCT 4351
Qy 3167 gaccctcctccagtacgtggatgacctgcttctggcgggagccaccaaacaggactgctt 3226
Db 4350 GATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCCACTTCTGAGCTTGACTGTCA 4291
Qy 3227 agaaggtacgaaggcactactgctggaattgtctgacctaggctacagagcctctgctaa 3286
Db 4290 ACAAGGTACGCGGGCCCTGTTACAAACCCTAGGGGACCTCGGATATCGGGCCTCGGCCAA 4231
Qy 3287 gaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggcgggca 3346
Db 4230 GAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCA 4171
Qy 3347 gcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaaccacagc 3406
Db 4170 GAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 4111
Qy 3407 caaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccggggtt 3466
Db 4110 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGATCCCTGGGTT 4051
Qy 3467 tgcgaccttagcagccccactctacccgctaaccaaagaaaaaggggggattctcctgggc 3526
Db 4050 TGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACGGGGACTCTGTTTGAGTGGGG 3991
Qy 3527 tcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctct 3586
Db 3990 CCCAGACCAGCAAAAGGCCTACCAAGAGATCAAGCAGGCTCTCTTAACTGCCCCTGCCCT 3931
Qy 3587 ggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaagggagtagc 3646
Db 3930 GGGATTGCCAGACTTGACTAAGCCCTTCGAACTTTTTGTTGACGAGAAGCAGGGCTACGC 3871
Qy 3647 ccgaggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctgtcaaa 3706
Db 3870 CAAAGGTGTCCTAACGCAAAAACTGGGGCCTTGGCGTCGGCCGGTGGCCTACCTGTCCAA 3811
Qy 3707 gaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggc 3766
Db 3810 AAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAGCAGCCATCGC 3751
Qy 3767 catactggtcaaggacgctgacaaattgactttgggacagaatataactgtaatagcccc 3826
Db 3750 CGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAGCCACTAGTCATTCTGGCCCC 3691
Qy 3827 ccatgcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgcccgcat 3886





```
RESULT  14
PCT-US93-08041-9/c
; Sequence 9, Application PC/TUS9308041
;  GENERAL INFORMATION:
;    APPLICANT:  The Public Health Research Institute of the City of
;    APPLICANT:  New York, Inc.
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
```

```
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10111
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/08041
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/938,100
;       FILING DATE:  28-AUG-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/_____
;       FILING DATE:  20-AUG-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Jacobs, Seth H.
;       REGISTRATION NUMBER:  32,140
;       REFERENCE/DOCKET NUMBER:  11698A50
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-757-2200
;       TELEFAX:  212-586-1461
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10367 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
PCT-US93-08041-9


  Query Match             19.4%;  Score 1596;  DB 5;  Length 10367;
  Best Local Similarity   58.4%;  Pred. No. 0;
  Matches 3161;  Conservative    0;  Mismatches 2135;  Indels  114;  Gaps   17;

Qy    467 tccgcggccgtccgactcttttgcctgcttgtggaa--gacgcggacgggtcgcgtgtgt 524
Db   7098 TCCGCCCCCGTCTGAGTTTTTGCTTTCGGTTTGGAACCGAAGCCGCGCCGCGCGTCTTGT 7039

Qy    525 ctggatctgttggtttctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagtttt 584
Db   7038 CTGCTGCAGCATCGTTCTGTGTTGTCTCTGTTTGACTGTT--TTTCTGTATTTGTCTGAA 6981

Qy    585 aatatgggacagacagtgactacccccctagtttgactctcgaccattggactgaagtt 644
Db   6980 AACATGGGCCAGGCTGTTACCACCCCCTTAAGTTTGACTTTAGACCACTGGAAGGATGTC 6921

Qy    645 agatccagggctcataatttgtcagttcaggttaagaagggaccttggcagactttctgt 704
Db   6920 GAACGGACAGCCCACAACCTGTCGGTAGAGGTTAGAAAAAGGCGCTGGGTTACATTCTGC 6861

Qy    705 gcctctgaatggccaacattcgatgttggatggccatcagaggggacctttaattctgaa 764
Db   6860 TCTGCAGAATGGCCAACCTTCAACGTCGGATGGCCACGAGACGGCACTTTTAACCCAGAC 6801

Qy    765 attatcctggctgttaaggcaatcatttttcagactggacccggctctcatcctgatcag 824
Db   6800 ATTATTACACAGGTTAAGATCAAGGTCTTCTCACCTGGCCCACATGGACATCCGGATCAG 6741

Qy    825 gagccctatatccttacgtggcaagatttggcagaagatcctccgccatgggttaaacca 884
Db   6740 GTCCCCTACATCGTGACCTGGGAAGCTATAGCAGTAGACCCCCCTCCCTGGGTCAGACCC 6681

Qy    885 tggctaaataaa----------ccaagaaagccaggtccccgaatcctggc 925
Db   6680 TTCGTGCACCCTAAACCTCCCCTCTCTCTTCCCCCTTCAGCCCCCTCTCTCCCACCTGAA 6621

Qy    926 tcttggagagaaaaacaaacactcggccgaaaaagtcgagccctcttcctcgtatctacc 985
Db   6620 CCCCCACTCTCGACCCCGCCCCAGTCCTCCCTCTATCCGGCTCTCACTTCTCCTTTAAAC 6561

Qy    986 ccgagatcgaggagccgccgacttggccggaaccccaacctgttcccccaccccttatc 1045
Db   6560 ACCAAACCTAGGCCTCAAGTCCTTCCTGATAGCGGAGGACCACTCATTGATCTACTCACG 6501

Qy   1046 cagcacagggtgctgtgaggggacctctgcccctcctggagctccggtggtggagggacc 1105
Db   6500 GAGGACCCTCCGCCTTACCGGGACCCAGGGCCACCCTCTCCTGACGGGAACGGCGATAGC 6441

Qy   1106 -------tgctgccgggactcggagccggagaggcgccaccccggagcggacagacgaga 1158
Db   6440 GGAGAAGTGGCCCCTACAGAAGGAGCCCCTGACCCTTCCCCAATGGTATCCCGCCTGCGG 6381

Qy   1159 tcgcgatattaccgctgcgcacctatggccctcccatgccaggggccaattgcagccc- 1217
Db   6380 GGAAGAAAAGAACCCCCCGTGGCGGATTCTACTACCTCTCAGGCGTTCCCCCTTCGCCTG 6321

Qy   1218 ---------------ctccagtattggccctttttcttctgcagatctctataattggaaa 1262
Db   6320 GGAGGGAATGGACAGTATCAATACTGGCCATTTTCCTCCTCTGACCTCTATAACTGGAAA 6261

Qy   1263 actaaccatcccccttttctcggaggatccccaacgcctcacggggttggtggagtccctt 1322
Db   6260 AATAACAACCCCCTCTTTCTCCGAGGACCCAGCTAAATTGACAGCTTTGATCGAGTCCGTT 6201

Qy   1323 atgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacactcttcaca 1382
Db   6200 CTCCTTACTCATCAGCCCACTTGGGATGACTGCCAACAGCTATTAGGGACCCTGCTGACG 6141

Qy   1383 accgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggccgacggg 1442
Db   6140 GGAGAAGAAAAACAGCGAGTGCTCCTAGAGGCCCGAAAGGCGGTTCGAGGGGAGGACGGA 6081

Qy   1443 cgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccccggttgg 1502
Db   6080 CGCCCAACTCAGCTGCCCAATGACATTAATGATGCTTTTCCCTTGGAACGTCCCGACTGG 6021

Qy   1503 gactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcg 1562
Db   6020 GACTACAACACCCAACGAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCG 5961

Qy   1563 ggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggtgatgcag 1622
Db   5960 GGTCTCCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGGATAACCCAG 5901

Qy   1623 ggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcaggcggttc 1682
Db   5900 GGACCTAATGAGTCTCCCTCAGCCTTTTTAGAGAGACTCAAGGAGGCCTATCGCAGATAC 5841

Qy   1683 acccctttgatcctacctcagaggcccagaaagcctcagtggccctggccttcattggg 1742
Db   5840 ACTCCTTATGACCCTGAGGACCCAGGGCAAGAAACCAATGTGGCCATGTCATTCATCTGG 5781

Qy   1743 cagtcggctctggatatcaggaagaaacttcagagactggaagggttacaggaggctgag 1802
Db   5780 CAGTCCGCCCCGGATATCGGGCGAAAGTTAGAGCGGTTAGAAGATTTGAAGAGTAAGACC 5721

Qy   1803 ttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacagaagaggag 1862
Db   5720 TTAGGAGACTTAGTGAGGGAAGCTGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAA 5661

Qy   1863 aaggaacagagaaaagaaaaggagagagaagaaagggaggaaagacgtgatagacggcaa 1922
Db   5660 AGAGAGGAACGTATTAGGAGAGAAACAGAGGAAAAGGAAGAACGCCGTAGGGCAGAGGAT 5601

Qy   1923 gagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcagggagaga 1982
Db   5600 GTGCAGAGAGAGAAGGAGAGGGACCGCAGAAGACATAGAGAAATGAGTA----------A 5551
```

```
Db   3393 CGGAGCAGCAGTAACCACCGAGACCGAGGTAGTCTGGGCCAAAGCACTGCCAGCCGGGAC 3334

Qy   4184 ttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaagggaa 4243
          || || ||||  |||||| | || || ||||| ||||| ||     |||| ||||| ||
Db   3333 ATCGGCCCAAAGAGCTGAGTTGATAGCGCTCACCCAAGCCTTAAAAATGGCAGAAGGTAA 3274

Qy   4244 atccataaacatttatacggacagcaggtatgcctttgcgactgcacacgtacacggggc 4303
            | ||  |||| || || ||| | ||||| ||||| ||||| ||  | ||||| |
Db   3273 GAAGCTGAATGTTTACACCGATAGCCGTTATGCTTTTGCCACTGCCCATATTCACGGAGA 3214

Qy   4304 catctataaacaaaggggttgcttacctcagcagggagggaaataaagaacaaagagga 4363
          || |||| |    | |||||||| || |||| ||| |  |||||| || || || || ||
Db   3213 AATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAAATCAAAAATAAGGACGA 3154

Qy   4364 aattctaagcctattagaagccttacatttgccaaaaaggctagctattatacactgtcc 4423
          ||  |  | |||| |  | ||  |      |||| ||||| ||    || || || || ||
Db   3153 GATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGACTTAGCATAATTCATTGCCC 3094

Qy   4424 tggacatcagaaagccaaagatctcatatctagagggaaccagatggctgaccgggttgc 4483
          ||||||||||| |  ||           | || || |||  |||||| ||||  |  ||
Db   3093 GGGACATCAGAAGGGAAACCGCGCGGAGGCAAGGGGCAACAGGATGGCCGACCAAGCGGC 3034

Qy   4484 caagcaggcagcccaggctgttaaccttctgcctataatagaaacgcccaaagccccaga 4543
          |    | | ||||      |  |  |    | ||   | |      |    ||    |||
Db   3033 CCGAGAAGTAGCCACTAGAGAAACTCCAGAGACTTCCACACTTCTGATAGAAAATTCAGC 2974

Qy   4544 accc-agacgacagtacaccctagaagactggcaagagataaaaaagatagaccagttct 4602
          ||| | ||    | |||   |       ||  |  || | |||||||   |    ||
Db   2973 CCCCTATACTCATGAACATTTTCACTATACGGTGACTGACATAAAAGATCTGACTAAACT 2914

Qy   4603 ctgagactccggaggggacc--------tgctatacctcatatgggaaggaaatcctgcc 4654
          | | | |    | ||          || |     |  || |||    |  ||||
Db   2913 AGGGGCCACTTATGACGATGCAAAGAAGTGTTGGGTTTATCAGGGAAAGCCTGTAATGCC 2854

Qy   4655 ccacaaagaagggttagaatatgtccaacagatacatcgtctaacccacctaggaactaa 4714
            |  ||      || |||    |  |     | ||||   | ||||||||  |
Db   2853 TGATCAATTCACCTTTGAACTATTAGATTTTCTTCATCAATTGACCCACCTCAGTTTCTC 2794

Qy   4715 acacctgcagcagttggtcagaaca------tccccttatcatgttctgaggctaccagg 4768
          | |     ||    | |   ||         |  |||||| |  | |||| |    |
Db   2793 AAAAACAAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTACATGCTGAACCGGGATCG 2734

Qy   4769 agtggctgactcggtggtcaaacattgtgtgccctgccagctggttaatgctaatccttc 4828
          |  |    |     |    |||     ||||    | ||| ||||| |     ||
Db   2733 AACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAGGTCAATGCCAGCAAGTC 2674

Qy   4829 cagaatacctccaggaaagagactaaggggaagccacccaggcgctcactgggaagtgga 4888
            |    | ||| |  ||| |  | ||   ||  || ||| | |||||||||  | ||
Db   2673 TGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCGACCCGGCACCCACTGGGAAATTGA 2614

Qy   4889 cttcactgaggtaaagccggctaaatacggaaacaaatatctattggtttttgtagacac 4948
          ||||||||||||| || |     || ||  | |||||||| || ||||| |||||||
Db   2613 TTTCACTGAGGTAAAACCTGGCCTGTATGGGTATAAATATCTTTTAGTTTTCATAGACAC 2554

Qy   4949 cttttcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtggctaa 5008
          || || |||||||||| |||| || || |||||||| ||| | |  || ||  | ||
Db   2553 TTTCTCTGGATGGGTAGAAGCTTTCCCAACCAAGAAAGAAACTGCCAAAGTTGTAACCAA 2494

Qy   5009 gaaaatactggaggaaatttttccaagatttggaatacctaaggtaatagggtcagacaa 5068
          |||  |||| || ||||| || || ||||| || || ||  ||||| | ||  | |||||
Db   2493 GAAGCTACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAGGTATTGGGAACCGACAA 2434

Qy   5069 tggtccagctttcgttgcccaggtaagtcagggactggccaagatattggggattgattg 5128
          ||| || || ||||  || |||||||||  | | ||| |  |||||||| |||||||
Db   2433 TGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGATTTATTGGGGGTTGATTG 2374

Qy   5129 gaaactgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaatagaac 5188
          |||||| |||||||| |||||||||| || ||||| |||||||| || |||||||| ||
Db   2373 GAAACTACATTGTGCTTACAGACCCCAGAGTTCAGGTCAGGTAGAAAGAATGAATAGGAC 2314

Qy   5189 cattaaagagacccttaccaaattgaccacagagactggcattaatgattggatggctct 5248
          || || |||||  | || ||||||||     | ||||||  ||  || ||| ||   ||
Db   2313 AATCAAGGAGACTTTAACTAAATTGACGCTTGCAACTGGCTCTAGGGACTGGGTGCTCCT 2254
```

```
Qy   5249 cctgccctttgtgcttttagggtgaggaacacccctggacagtttgggctgacccccta 5308
          || ||| |||  || | | | |    | ||||| || || |     || || ||||| ||
Db   2253 GCTTCCCCTTGCCCTGTATCGAGCCCGCAACACGCCGGGCCCCCACGGTCTCACCCCATA 2194

Qy   5309 tgaattgctctacgggggaccccccccgttggcagaaattgcctttgcacatagtgctga 5368
          |||| |  | || |||| |||||| ||  | | | |  |  | |  | |  || |
Db   2193 TGAAATCTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGATCCTGACATGGCAAA 2134

Qy   5369 tgtgctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcagcg 5428
          ||    | |  ||   || |     ||     |  |||| ||| |   |||  |||
Db   2133 GGTTACTCATAACCCCTCTCTCCAAGCTCATTTACAGGCACTCTACCTGGTCCAGCACGA 2074

Qy   5429 agcgtggaagcagctccgggaggcctactcag------gaggagacttgcaagttccaca 5482
          ||  ||||  | | |   ||  || |||  ||       ||       |  ||| || ||
Db   2073 AGTCTGGAGACCGTTGGCGGCAGCTTACCAAGAGCAACTAGACCGGCCGGTAGTACCTCA 2014

Qy   5483 tcgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaaacctcgagac 5542
            |  ||||  || || ||  |||| |  ||  |  | ||||   |  ||| || ||  |
Db   2013 CCCTTTCCGGGTCGGTGACACAGTGTGGGTCCGCAGACACCAAACTAAAAATCTAGAACC 1954

Qy   5543 tcggtggaagggaccttatctcgtacttttgaccacaccaacggctgtgaaagtcgaagg 5602
          || ||||| ||||| ||| ||| ||  |||| || || || ||| | ||||| || ||
Db   1953 CCGCTGGAAAGGACCCTATACCGTCCTACTGACTACCCCCACCGCTCTCAAAGTAGACGG 1894

Qy   5603 aatccccacctggatccatgcatcccacgttaagccggc---------------gccacc 5647
          ||  |  | ||||||||  ||  |||||||  ||| | ||               ||||||
Db   1893 CATTGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGACACCAGGATTGAGCCACC 1834

Qy   5648 tcccgactcggggtggagagccgaaaagactgagaatccccttaagcttcgcctccatcg 5707
            | || |||   ||| | |   ||     |  | |||||||| ||| |  |  |    ||
Db   1833 AGCAGAATCGACATGGCGTGTTCAACGCTCCCAAAATCCCCTAAAGATAAGATTGACCCG 1774

Qy   5708 cctggttccttactctaacaataactccccaggccagtagtaaacgccttatagacagct 5767
          |  |    | || ||    |  |||  |  | |  |    |||  ||   ||||
Db   1773 CGGGACCTCCTAATCCCCTTAATTCTCTTCCTGTCTCTCAAAGGGGCCAGATCCGCAGCA 1714

Qy   5768 cgaaccccca 5777
          |    | |||
Db   1713 CCCGGCTCCA 1704
```

```
RESULT  15
US-09-111-085-1
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
;  APPLICANT: Stoye, Jonathan P
;  APPLICANT: Weiss, Robin A
;  TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
;  TITLE OF INVENTION: specific sequences
;  FILE REFERENCE: 4238/75168
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2462
;   TYPE: DNA
;   ORGANISM: Porcine retrovirus
US-09-111-085-1


  Query Match           19.1%;  Score 1568.4;  DB 3;  Length 2462;
  Best Local Similarity   79.6%;  Pred. No. 0;
  Matches 2002;  Conservative    0;  Mismatches  396;  Indels  116;  Gaps    8;

Qy   5410 tcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagact 5469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 tcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagact 60
```

```
Db   2169 gagccaaggagaaactgacctctagccttcccagttctaagattagaactattaacaaga 2228

Qy   7626 caagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagtt 7685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2229 caagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagtt 2288

Qy   7686 aataaaaagctctaaatgcccccgaattacagaccctgctggctgccagtaaataggtag 7745
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   2289 aataaaaagctctaaatgcccccgaattccagaccctgctggctgccagtaaataggtag 2348

Qy   7746 aaggtcacacttcctattgttccagggcctgctatcctggcctaagtaagataacaggaa 7805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2349 aaggtcacacttcctattgttccagggcctgctatcctggcctaagtaagataacaggaa 2408

Qy   7806 atgagttgactaatcgcttatctggattctgtaaaactgactggcaccatagaa 7859
          |||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   2409 atgagttgactaatcgcttatctggattctgtaaaaccgactggcaccatagaa 2462
```

Search completed: February 24, 2002, 02:58:34
Job time: 21265 sec

```
OM of: US-09-171-553B-4 to: Issued_Patents_NA:*   out_format : pfs

Date: Feb 24, 2002 10:12 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09171553/runat_22022002_122826_11171/app_query.fasta_1.3312
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=feb22std.rni
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09171553_@CGN1_1_158 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-171-553B-4
Query length: 524
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 134.760000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-2 + 2758.00 3838.83  3.6e-206  7333  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-3 + 2640.00 3671.81  7.2e-197  8132  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-1 + 2634.00 3663.48  2.1e-196  8060  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-075-272-1 + 1684.00 2327.30  5.5e-122  8655  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-1 + 1660.50 2294.43  3.8e-120  8535  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-3 + 1355.50 1867.43  2.3e-96   7308  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-4 + 1355.50 1867.43  2.3e-96   7308  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-2 + 1355.50 1866.97  2.4e-96   7616  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-850-961-1 + 1355.50 1865.97  2.8e-96   8332  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-13 - 1354.50 1864.74  3.2e-96   8202  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-8 + 1350.50 1858.95  6.8e-96   8323  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-8 + 1350.50 1858.95  6.8e-96   8323  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-8 + 1350.50 1858.95  6.8e-96   8323
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-9 - 1350.50 1856.50  9.3e-96   10367  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-9 - 1350.50 1856.50  9.3e-96   10367  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-9 - 1350.50 1856.50  9.3e-96   10367
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-105-483-324 + 1344.00 1858.92  6.8e-96   3674  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-709-209-324 + 1344.00 1858.92  6.8e-96   3674  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-101-324 + 1344.00 1858.92  6.8e-96   3674  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-181-335B-3 +  822.00 1142.31  5.6e-56   789  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-00129-3 +  822.00 1142.31  5.6e-56   789
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-181-335B-6 +  819.00 1138.09  9.6e-56   789  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-00129-6 +  819.00 1138.09  9.6e-56   789
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-181-335B-5 +  816.00 1133.87  1.7e-55   789  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-00129-5 +  816.00 1133.87  1.7e-55   789
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-532-814-1 +  758.50 1026.71  1.5e-49   8387  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-991-789A-141 +  640.00  858.88  3.4e-40   9388
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-991-789A-227 +  636.00  863.80  1.8e-40   3646
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-3 +  560.00  746.11  6.6e-34   9661  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-654-737B-3 +  421.50  556.99  2.2e-23   5865  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-793-610-5 +  421.50  555.83  2.6e-23   6505  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-793-610-6 +  413.50  540.58  1.8e-22   9318  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-753-520B-4 +  411.00  538.34  2.5e-22   8316  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-753-520B-3 +  411.00  537.32  2.8e-22   9115  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-352-990-1 +  406.00  534.29  4.1e-22   6365  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-793-610-3 +  404.00  533.53  4.5e-22   5292  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-793-610-2 +  404.00  533.53  4.6e-22   5294  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-793-610-1 +  404.00  533.47  4.6e-22   5323  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-793-610-4 +  404.00  533.38  4.6e-22   5364  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-336-132-1 +  403.00  530.46  6.7e-22   6145  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-935-312-1 +  403.00  530.46  6.7e-22   6145
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-848-760B-1 +  403.00  530.46  6.7e-22   614
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-789-333F-58 +  402.00  531.47  5.9e-22   49
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-133-944-1 -  401.50  522.82  1.8e-21   1010
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-182-612B-3 +  399.50  556.18  2.5e-23   393

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-2

seq_documentation_block:
; Sequence 2, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-2


alignment_scores:
            Quality: 2758.00              Length:    525
              Ratio:  5.263                Gaps:      1
 Percent Similarity:  99.810   Percent Identity: 99.619

alignment_block:
 US-09-171-553B-4 x US-08-766-528-2   ..

 Align seg 1/1  to: US-08-766-528-2  from: 1  to: 7333

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     598 ATGGGACAGACAGTGACTACCCCCCTTAGTTTGACTCTCGACCATTGGAC 647

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     648 TGAAGTTAGATCCAGGGCTCATAATTTGTCAGTTCAGGTTAAGAAGGGAC 697

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     698 CTTGGCAGACTTTCTGTGCCTCTGAATGGCCAACATTCGATGTTGGATGG 747
```

```
 51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
748 CCATCAGAGGGGACCTTTAATTCTGAAATTATCCTGGCTGTTAAGGCAAT 797

 67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
798 CATTTTTCAGACTGGACCCGGCTCTCATCCTGATCAGGAGCCCTATATCC 847

 84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
848 TTACGTGGCAAGATTTGGCAGAAGATCCTCCGCCATGGGTTAAACCATGG 897

101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||
898 CTAAATAAACCAAGAAAGCCAGGTCCCCGAATCCTGGCTCTTGGAGAGAA 947

117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
    |||||||||||||||||||||||||||||||   ||||||||||||||||
948 AAACAAACACTCGGCCGAAAAAGTCGAGCCCTC.TCCTCGTATCTACCCC 996

134 rgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
997 GAGATCGAGGAGCCGCCGACTTGGCCGGAACCCCAACCTGTTCCCCCACC 1046

151 ProLeuSerSerThrGlyCysCysGluGlyThr.SerAlaProProGlyA 167
    |||||||||||||||||||||||||||||||| ||||||||||||||||
1047 CCCTTATCCAGCACAGGGTGCTGTGAGGGGACCCTCTGCCCCTCCTGGAG 1096

167 laProValValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAla 183
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1097 CTCCGGTGGTGGAGGGACCTGCTGCCGGGACTCGGAGCCGGAGAGGCGCC 1146

184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGl 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1147 ACCCCGGAGCGGACAGACGAGATCGCGATATTACCGCTGCGCACCTATGG 1196

200 yProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheS 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1197 CCCTCCCATGCCAGGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTT 1246

217 erSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1247 CTTCTGCAGATCTCTATAATTGGAAAACTAACCATCCCCCTTTCTCGGAG 1296

234 AspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGl 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1297 GATCCCCAACGCCTCACGGGGTTGGTGGAGTCCCTTATGTTCTCTCACCA 1346

250 nProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrThrG 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1347 GCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTCACAACCG 1396

267 luGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAla 283
    ||||||||||||||||||||||||||||||:::||||||||||||||||||
1397 AGGAGCGAGAGAGAATTCTGTTAGAGGCTAAAAAAAAATGTTCCTGGGGCC 1446

284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLe 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1447 GACGGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGATTTCCCTT 1496

300 uThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuL 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1497 GACTCGCCCCGGTTGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGA 1546

317 ysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1547 AAATCTATCGCCAGGCTCTGGTGGCGGGTCTCCGGGGCGCCTCAAGACGG 1596

334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluPr 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1597 CCCACTAATTTGGCTAAGGTAAGAGAGGTGATGCAGGGACCGAACGAACC 1646

350 oProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrP 367
```

```
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1647 TCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTCAGGCGGTTCACCC 1696

 367 roPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPhe 383
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1697 CTTTTGATCCTACCTCAGAGGCCCAGAAAGCCTCAGTGGCCCTGGCCTTC 1746

 384 IleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGl 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1747 ATTGGGCAGTCGGCTCTGGATATCAGGAAGAAACTTCAGAGACTGGAAGG 1796

 400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValT 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1797 GTTACAGGAGGCTGAGTTACGTGATCTAGTGAGAGAGGCAGAGAAGGTGT 1846

 417 yrTyrArgArgGluThrGluGluGluLysGluGlnArgLysGluLysGlu 433
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1847 ATTACAGAAGGGAGACAGAAGAGGAGAAGGAACAGAGAAAAGAAAAGGAG 1896

 434 ArgGluGluArgGluGluArgArgAspArgArgGlnGluLysAsnLeuTh 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1897 AGAGAAGAAAGGGAGGAAAGACGTGATAGACGGCAAGAGAAGAATTTGAC 1946

 450 rLysIleLeuAlaAlaValValGluGlyLysSerSerArgGluArgGluA 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1947 TAAGATCTTGGCCGCAGTGGTTGAAGGGAAGAGCAGCAGGGAGAGAGAGA 1996

 467 rgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuGly 483
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1997 GAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGC 2046

 484 AsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGl 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2047 AATAGGACCCCACTCGACAAGGACCAGTGTGCGTATTGTAAAGAAAAAGG 2096

 500 yHisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValL 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2097 ACACTGGGCAAGGAACTGCCCCAAGAAGGGAAACAAAGGACCGAAGGTCC 2146

 517 euAlaLeuGluGluAspLysAsp 524
     |||||||||||||||||||||||
2147 TAGCTCTAGAAGAAGATAAAGAT 2169
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-3

```
seq_documentation_block:
; Sequence 3, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
```

```
;       REGISTRATION NUMBER:  35,965
;       REFERENCE/DOCKET NUMBER:  MGP-038CP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227-5941
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8132 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-766-528-3


alignment_scores:
            Quality: 2640.00              Length:    525
              Ratio:  5.116                Gaps:      1
 Percent Similarity:  98.286   Percent Identity: 94.476

alignment_block:
 US-09-171-553B-4 x US-08-766-528-3   ..

 Align seg 1/1  to: US-08-766-528-3  from: 1  to: 8132

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     585 ATGGGACAGACGGTGACGACCCCTCTTAGTTTGACTCTCGACCATTGGAC 634

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
         |||||||:::|||||||||||||||||||||||||||||||||||||||
     635 TGAAGTTAAATCCAGGGCTCATAATTTGTCAGTTCAGGTTAAGAAGGGAC 684

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
         ||||||||||||||||:::|||||||||||||||||||||||||||||||
     685 CTTGGCAGACTTTCTGTGTCTCTGAATGGCCGACATTCGATGTTGGATGG 734

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||::
     735 CCATCAGAGGGGACCTTTAATTCTGAGATTATCCTGGCTGTTAAAGCAGT 784

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
         :|||||||||||||||||||||||||||||||||||||||||||||||||
     785 TATTTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAGCCCTATATCC 834

      84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     835 TTACGTGGCAAGATTTGGCAGAGGATCCTCCGCCATGGGTTAAACCATGG 884

     101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     885 CTGAATAAGCCAAGAAAGCCAGGTCCCCGAATTCTGGCTCTTGGAGAGAA 934

     117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
         |||||||||||||||||||||||||:::|||   ||||||||||||||||
     935 AAACAAACACTCGGCTGAAAAAGTCAAGCCCTC.TCCTCATATCTACCCC 983

     134 rgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
         |||||   |||||||:::   ||||||||||||||   ||||||||||||
     984 GAGATTGAGGAGCCACCGGCTTGGCCGGAACCCCAATCTGTTCCCCCACC 1033

     151 ProLeuSerSerThrGlyCysCysGluGlyThr.SerAlaProProGlyA 167
         |||||||||:::|||||||||   |||||||||    ||||||||||||
    1034 CCCTTATCTGGCACAGGGTGCCGCGAGGGGACCCTTTGCCCCTCCTGGAG 1083

     167 laProValValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAla 183
         |||||:::||||||||||||||||||||||||||||||||||||||||||
    1084 CTCCGGCGGTGGAGGGACCTGCTGCAGGGACTCGGAGCCGGAGGGGCGCC 1133

     184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGl 200
         ||||||||||||||||||||||||   ||||||||||||||||||||||
    1134 ACCCCGGAGCGGACAGACGAGATCGCGACATTACCGCTGCGCACGTACGG 1183

     200 yProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheS 217
         |||||||   ||||||||||||||||||||||||||||||||||||||||
    1184 CCCTCCCACACCGGGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTT 1233

     217 erSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1234 CTTCTGCAGATCTCTATAATTGGAAAACTAACCATCCCCCTTTCTCGGAG 1283

     234 AspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGl 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1284 GATCCCCAACGCCTCACGGGGTTGGTGGAGTCCCTTATGTTCTCTCACCA 1333

     250 nProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrThrG 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1334 GCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTCACAACCG 1383

     267 luGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAla 283
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1384 AGGAGCGAGAGAGAATTCTATTAGAGGCTAGAAAAAATGTTCCTGGGGCC 1433

     284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLe 300
         ||||||||||||||||:::|||||||||||||||||||||||||||||||
    1434 GACGGGCGACCCACGCGGTTGCAAAATGAGATTGACATGGGATTTCCCTT 1483

     300 uThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuL 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1484 AACTCGCCCCGGTTGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGA 1533

     317 ysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1534 AAATCTATCGCCAGGCTCTGGTGGCGGGTCTCCGGGGCGCCTCAAGACGG 1583

     334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluPr 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1584 CCCACTAATTTGGCTAAGGTAAGAGAAGTGATGCAGGGACCGAATGAACC 1633

     350 oProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrP 367
         ||||||||||||||||||||||||:::|||||||||||||||:::||||
    1634 CCCCTCTGTTTTTCTTGAGAGGCTCTTGGAAGCCTTCAGGCGGTACACCC 1683

     367 roPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPhe 383
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1684 CTTTTGATCCCACCTCAGAGGCCCAAAAAGCCTCAGTGGCTTTGGCCTTT 1733

     384 IleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGl 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1734 ATAGGACAGTCAGCCTTGGATATTAGAAAGAAGCTTCAGAGACTGGAAGG 1783

     400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValT 417
         ||||||||||||||||||||||||||||||||:::|||||||||||||||
    1784 GTTACAGGAGGCTGAGTTACGTGATCTAGTGAAGGAGGCAGAGAAAGTAT 1833

     417 yrTyrArgArgGluThrGluGluGluLysGluGlnArgLysGluLysGlu 433
         |||||:::|||||||||||||||||||:::|||||||||||||||:::|||
    1834 ATTACAAAAGGGAGACAGAAGAAGAAAGGGAACAAAGAAAAGAGAGAAA 1883

     434 ArgGluGluArgGluGluArgArgAspArgArgGlnGluLysAsnLeuTh 450
         ||||||||||||||||||||||||||:::::|||||||||||||||||||
    1884 AGAGAGGAAAGGGAGGAAAGACGTAATAAACGGCAAGAGAAGAATTTGAC 1933

     450 rLysIleLeuAlaAlaValValGluGlyLysSerSerArgGluArgGluA 467
         ||||||||||||||||||||||||||||||||:::   ||||||||||
    1934 TAAGATCTTGGCTGCAGTGGTTGAAGGGAAAAGCAATACGGAAAGAGAGA 1983

     467 rgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuGly 483
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1984 GAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGC 2033

     484 AsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGl 500
         ||||||||||||||||||||||||||||||||||||||||||||||:::||
    2034 AATAGGACCCCACTCGACAAGGACCAATGTGCATATTGTAAAGAAAGAGG 2083

     500 yHisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValL 517
```

```
      |||||||||||||||||||||||||||||||||||||||||||||||:::::::|
   2084 ACACTGGGCAAGGAACTGCCCCAAGAAGGGAAACAAAGGACCAAGGATCC 2133

    517 euAlaLeuGluGluAspLysAsp 524
        |||||||||||||||||||||||
   2134 TAGCTCTAGAAGAAGATAAAGAT 2156

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-1

seq_documentation_block:
; Sequence 1, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;     APPLICANT:  Jay A. Fishman
;     TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;     TITLE OF INVENTION:  AND METHODS OF USE
;     NUMBER OF SEQUENCES:  74
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;       STREET:  60 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109-1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/766,528
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/572,645
;       FILING DATE:  14-DEC-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Louis Myers
;       REGISTRATION NUMBER:  35,965
;       REFERENCE/DOCKET NUMBER:  MGP-038CP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8060 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-766-528-1


alignment_scores:
            Quality: 2634.00              Length:    525
              Ratio:  5.105                Gaps:      1
 Percent Similarity:  98.286   Percent Identity: 94.286

alignment_block:
 US-09-171-553B-4 x US-08-766-528-1   ..

 Align seg 1/1  to: US-08-766-528-1  from: 1  to: 8060

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3112 ATGGGACAGACGGTGACGACCCCTCTTAGTTTGACTCTCGACCATTGGAC 3161

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
         |||||||:::|||||||||||||||||||||||||||||||||||||||||
    3162 TGAAGTTAAATCCAGGGCTCATAATTTGTCAGTTCAGGTTAAGAAGGGAC 3211

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
         |||||||||||||||||:::|||||||||||||||||||||||||||||||
    3212 CTTGGCAGACTTTCTGTGTCTCTGAATGGCCGACATTCGATGTTGGATGG 3261

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3262 CCATCAGAGGGGACCTTTAATTCTGAGATTATCCTGGCTGTTAAAGCAAT 3311

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3312 TATTTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAGCCCTATATCC 3361

      84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3362 TTACGTGGCAAGATTTGGCAGAGGATCCTCCGCCATGGGTTAAACCATGG 3411

     101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3412 CTGAATAAGCCAAGAAAGCCAGGTCCCCGAATTCTGGCTCTTGGAGAGAA 3461

     117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
         |||||||||||||||||||||||||:::|||   ||||||||||||||||
    3462 AAACAAACACTCGGCTGAAAAAGTCAAGCCCTC.TCCTCATATCTACCCC 3510

     134 rgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
         |||||   |||::::::   ||||||||||||||||   ||||||||||||
    3511 GAGATTGAGGAACCACCGGCTTGGCCGGAACCCCAATCTGTTCCCCCACC 3560

     151 ProLeuSerSerThrGlyCysCysGluGlyThr.SerAlaProProGlyA 167
         |||||||||:::|||||||||  |||||||||    ||||||||||||
    3561 CCCTTATCTGGCACAGGGTGCCGCGAGGGGACCCTTTGCCCCTCCTGGAG 3610

     167 laProValValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAla 183
         |||||:::|||||||||||||:::|||||||||||||||||||||||||||
    3611 CTCCGGCGGTGGAGGGACCTTCTGCAGGGACTCGGAGCCGGAGGGGCGCC 3660

     184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGl 200
         |||||||||||||||||||||||||||   ||||||||||||||||||||
    3661 ACCCCGGAGCGGACAGACGAGATCGCGACATTACCGCTGCGCACGTACGG 3710

     200 yProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheS 217
         |||||||   |||||||||||||||||||||||||||||||||||||||||
    3711 CCCTCCCACACCGGGGGGCCAATTGCAGCCCCTCCAGTATTGGCCCTTTT 3760

     217 erSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3761 CTTCTGCAGATCTCTATAATTGGAAAACTAACCATCCCCCTTTCTCGGAG 3810

     234 AspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGl 250
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3811 GATCCCCAACGCCTCACGGGGTTGGTGGAGTCCCTTATGTTCTCTCACCA 3860

     250 nProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrThrG 267
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3861 GCCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTCACAACCG 3910

     267 luGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAla 283
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3911 AGGAGCGAGAGAGAATTCTATTAGAGGCTAGAAAAAATGTTCCTGGGGCC 3960

     284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLe 300
         |||||||||||||||||:::|||||||||||||||||||||||||||||||
    3961 GACGGGCGACCCACGCGGTTGCAAAATGAGATTGACATGGGATTTCCCTT 4010

     300 uThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuL 317
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    4011 AACTCGCCCCGGTTGGGACTACAACACGGCTGAAGGTAGGGAGAGCTTGA 4060

     317 ysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    4061 AAATCTATCGCCAGGCTCTGGTGGCGGGTCTCCGGGGCGCCTCAAGACGG 4110

     334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluPr 350
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    4111 CCCACTAATTTGGCTAAGGTAAGAGAAGTGATGCAGGGACCGAATGAACC 4160
```

```
      350 oProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrP 367
          ||||||||||||||||||||||||||:::|||||||||||||||:::||||
     4161 CCCCTCTGTTTTTCTTGAGAGGCTCTTGGAAGCCTTCAGGCGGTACACCC 4210

      367 roPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPhe 383
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     4211 CTTTTGATCCCACCTCAGAGGCCCAAAAAGCCTCAGTGGCTTTGGCCTTT 4260

      384 IleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGl 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     4261 ATAGGACAGTCAGCCTTGGATATTAGAAAGAAGCTTCAGAGACTGGAAGG 4310

      400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValT 417
          ||||||||||||||||||||||||||||||||:::||||||||||||||||
     4311 GTTACAGGAGGCTGAGTTACGTGATCTAGTGAAGGAGGCAGAGAAAGTAT 4360

      417 yrTyrArgArgGluThrGluGluGluLysGluGlnArgLysGluLysGlu 433
          |||||:::|||||||||||||||||:::|||||||||||||||:::|||
     4361 ATTACAAAAGGGAGACAGAAGAAGAAAGGGAACAAAGAAAAGAGAGAGAA 4410

      434 ArgGluGluArgGluGluArgArgAspArgArgGlnGluLysAsnLeuTh 450
          ||||||||||||||||||||||::::::||||||||||||||||||||||
     4411 AGAGAGGAAAGGGAGGAAAGACGTAATAAACGGCAAGAGAAGAATTTGAC 4460

      450 rLysIleLeuAlaAlaValValGluGlyLysSerSerArgGluArgGluA 467
          ||||||||||||||||||||||||||||||||:::   ||||||||||
     4461 TAAGATCTTGGCTGCAGTGGTTGAAGGGAAAAGCAATACGGAAAGAGAGA 4510

      467 rgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuGly 483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     4511 GAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGC 4560

      484 AsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGl 500
          ||||||||||||||||||||||||||||||||||||||||||||||||:::||
     4561 AATAGGACCCCACTCGACAAGGACCAATGTGCATATTGTAAAGAAAGAGG 4610

      500 yHisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValL 517
          ||||||||||||||||||||||||||||||||||||||||||||||:::::|
     4611 ACACTGGGCAAGGAACTGCCCCAAGAAGGGAAACAAAGGACCAAGGATCC 4660

      517 euAlaLeuGluGluAspLysAsp 524
          |||||||||||||||||||||||
     4661 TAGCTCTAGAAGAAGATAAAGAT 4683

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-075-272-1

seq_documentation_block:
; Sequence 1, Application US/09075272
; Patent No. 6136598
;  GENERAL INFORMATION:
;    APPLICANT:  MILLER, A. DUSTY
;    APPLICANT:  WOLGAMOT, GREG
;    APPLICANT:  BONHAM, LYNN
;    TITLE OF INVENTION:  MUS DUNNI ENDOGENOUS RETROVIRAL
;    TITLE OF INVENTION:  PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,272
;      FILING DATE:  08-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/046,140
;      FILING DATE:  09-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poor, Brian W.
;      REGISTRATION NUMBER:  32,928
;      REFERENCE/DOCKET NUMBER:  14538A-003710
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-075-272-1


alignment_scores:
           Quality: 1684.00              Length:    541
             Ratio:  3.907                Gaps:     10
 Percent Similarity:  79.667   Percent Identity: 61.553

alignment_block:
 US-09-171-553B-4 x US-09-075-272-1   ..

 Align seg 1/1  to: US-09-075-272-1  from: 1  to: 8655

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         |||||||||:::|||:::|||||||||||||||||||||||:::||||||
     722 ATGGGACAATCTGTCTCCACTCCCCTTTCTCTAACCCTGGAGCATTGGAA 771

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
         ||||||:::   ||||||||||||   ||||||:::|||:::|||||||
     772 GGAGGTAAAAATCAGAGCACACAATCAGTCAGTGGAGGTTAGGAAAGGCC 821

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
         ||||||||||||||||||||||||||||||||||   ||||||||
     822 CATGGCAAACTTTTTGCGCCTCCGAGTGGCCAACGTTTGGAGTGGGCTGG 871

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||   ||||||:::|||:::   :::::|||   ||||||:::   ||
     872 CCACCGGAGGGTGCTTTTGACTTGTCACTGATCGCCGCCGTCAGGCGAAT 921

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
         |:::|||||         |||:::||||||||||   ||||||||:
     922 TGTTTTCAGGAA...GAAGGGGGTCACCCTGATCAGATCCCCTACATTG 968

      84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
         ::|||||||||:::|||:::::   ||||||   ||||||||||||||||
     969 TGACCTGGCAGAGTCTCGTCCAGTTCCCACCTTCGTGGGTCAAGCCCTGG 1018

     101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
                        |||:::|||      |||:::::::::
    1019 ..............ACCCCAAATCCTTCGAAACTGACGGTTGCAGTTGC 1053

     117 sAsnLysHisSerAlaGluLysValGluProSerSer......SerTyrL 132
          ::::::   :::|||||||||      ||||||:::      :::
    1054 CCAGTCCGATGCAGCTGAGAAATCTGGTCCGTCAGCACCCCCCAAGATTT 1103

     132 euProArgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSer 148
           |||:::                  :::   :::   |||
    1104 ATCCAGAGATTGACGACCTCCTCTGGATGGACTCCCAACCTCCCCCTTAC 1153

     149 ProThrProLeuSerSerThrGlyCysCysGluGlyThrSerAlaPro.. 164
         |||   |||   :::     :::::::::     ::::::||||||
    1154 CCCCTGCCCCAGCAGCCACCTGCAGCAGCCCCACCTGTGGCAGCCCCTCA 1203

     165 .......ProGlyAlaProValValGluGlyProAlaAlaGlyThrArgS 179
                 |||   |||      ::::::||||||||:::||||||||||
    1204 GCCGGAACCAACAGCAAGCGGGGCTCAGGGACCGGCGGGAGGGACTCGGA 1253
```

```
 179 erArgArgGlyAlaThrProGluArg.........ThrAspGluIleAla 192
     ||||||||||||   :::|||   :::            |||:::   :::
1254 GCCGCCGGGGTCGGAGCCCTGCAGAGGAAGGGGGGCCAGATTCCACAGTT 1303

 193 IleLeuProLeuArgThrTyr...GlyProProMetProGly...GlyGl 207
     ||||||||||||:::::::   |||   |||   ||||||   :::::
1304 GCCTTGCCCCTTAGAGCTCATGTGGGAGGGCCAACGCCAGGACCTAATGA 1353

 207 nLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAspLeuTyrAsnT 224
     :|||   |||||||||||||||||||||||||:::||||||||||||||
1354 TCTCATTCCTTTACAGTACTGGCCTTTTTCTTCTTCTGATTTATATAATT 1403
                ↑
 224 rpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240
     ||||||||||||||||||||||||||||:::|||:::   ||||||||
1404 GGAAAACTAACCACCCTCCCTTCTCAGAAAACCCCTCTGGGCTTACTGGG 1453

 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGl 257
     |||:::||||||||||||||||||||||||||||||||||||||||||||
1454 CTCCTTGAATCACTTATGTTCTCTCATCAGCCCACTTGGGATGATTGTCA 1503

 257 nGlnLeuLeuGlnThrLeuPheThrThrGluGluArgGluArgIleLeuL 274
     ||||||||||||:::|||||||||||||||||||||||||||||||||:
1504 GCAGCTTTTGCAGGTTCTCTTTACCACAGAGGAAAGGGAAAGAATCCTGA 1553

 274 euGluAlaArgLysAsnValProGlyAlaAspGlyArgProThrGlnLeu 290
     ::|||||||||||||||||   |||   ||||||   ||||||   |||
1554 TGGAGGCAAGAAAAAACGTTCTAGGAGAGGACGGCACACCCACTGCTCTC 1603

 291 GlnAsnGluIleAspMetGlyPheProLeuThrArgProGlyTrpAspTy 307
     |||   :::|||   :::|||||||||:::||||||:::||||||||
1604 CCTAACCTCGTGGACGAGGCTTTCCCCTTGAACCGCCCCAACTGGGACTA 1653

 307 rAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArgGlnAlaLeuV 324
     ||||||||||||||||||      |||   :::||||||:::::||||
1654 CAACACTGCGGAAGGTAGGGGACGCCTCCTTGTCTACCGTCGGACTCTGG 1703

 324 alAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340
     ||||||||||||||||||:::||||||||||||||||||||||||||||
1704 TGGCAGGTCTCAGAGGAGCCGCAAGACGGCCCACCAATTTGGCTAAGGTA 1753

 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluAr 357
     |||||||||:::||||||   :::|||||||||||||||||||||||||
1754 AGAGAGGTCTTGCAGGGGCAGACTGAACCACCCTCAGTCTTTCTGGAGCG 1803

 357 gLeuMetGluAlaPheArgArgPheThrProPheAspProThrSerGluA 374
     ||||||||||||:::||||||:::|||||||||||||||:::||||||:
1804 TCTTATGGAGGCGTATAGGAGGTACACCCCTTTCGATCCCTCGTCAGAGG 1853

 374 laGlnLysAlaSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAsp 390
     ::||||||||:::||||||:::|||||||||||||||||||   |||
1854 GACAGAAAGCCGCTGTAGCCATGGCCTTCATTGGCCAGTCTGCTCCCGAT 1903

 391 IleArgLysLysLeuGlnArgLeuGluGlyLeuGlnGluAlaGluLeuAr 407
     |||:::||||||||||||||||||||||||||:::      |||::
1904 ATTAAGAAAAAGCTACAGAGGCTGGAGGGGCTCCAAGATTATACGCTCCA 1953

 407 gAspLeuValArgGluAlaGluLysValTyrTyrArgArgGluThrGluG 424
     :|||||||||:::|||||||||||||||||:::::::|||||||||||||
1954 AGATTTAGTGAAGGAAGCAGAGAAGGTTTATCACAAGAGAGAAACAGAAG 2003

 424 luGluLysGluGlnArgLysGluLysGluArgGluGluArgGluGluArg 440
     |||||:::::::::|||:::::|||||   ||||||||||||:::|||
2004 AAGAGAGGCAGGAGAGAGAAAAGAAAGAAGTAGAGGAGAGGGAAAATAGG 2053

 441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValVa 457
     ||||||||||||||||||:::|||||||:::||||||||||||||||::
2054 CGAGATCGCCGTCAGGAGAGAAATTTGAGTAAGATTCTGGCCGCAGTTAT 2103

 457 lGluGlyLysSerSerArgGluArgGluArgAspPheArgLysIleArgS 474
     ::::   ::::::|||::::::
2104 AAATGATAGGCAGTCAGAAAAA..................................... 2125

 474 erGlyProArgGlnSerGlyAsnLeuGlyAsnArg..... 485
```

```
                 ::::::|||   ||||||||||||
    2126 ........GGCAGAACAGGATTCCTGGGCAACAGGGCAGTGAAACCGCCA 2167

     486 ............ThrProLeuAspLysAspGlnCysAlaTyrCysLysGl 498
                     |||||||||:::|||||||||||||||:::|||||||||
    2168 GGTGGCAGAAAGACGCCGCTGGAAAAAGACCAATGCGCCTTTTGCAAAGA 2217

     498 uLysGlyHisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProL 515
         ||||||||||||||||:::::::||||||||||   ::::::      |
    2218 AAAAGGACACTGGGCTAAAGACTGCCCTAAGAAA...AGAAGGCAATTCA 2264

     515 ysValLeuAlaLeuGluGluAsp 522
         |||||||||:::|||||||:::|||
    2265 AGGTCCTGACCCTAGAAGACGAT 2287

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-1

seq_documentation_block:
; Sequence 1, Application US/08716351A
; Patent No. 6033905
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Gibbon Ape Leukemia Virus-Based
;    TITLE OF INVENTION:  Retroviral Vectors
;    NUMBER OF SEQUENCES:  5
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/716,351A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/03784
;      FILING DATE:  06-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian,, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  15280-128-1PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..8535
;      OTHER INFORMATION:  /standard_name= "GaLV SEATO Genome"
US-08-716-351A-1


alignment_scores:
            Quality: 1660.50            Length:    542
              Ratio:  3.944              Gaps:     10
 Percent Similarity:  77.675   Percent Identity: 61.439

alignment_block:
 US-09-171-553B-4 x US-08-716-351A-1   ..

 Align seg 1/1  to: US-08-716-351A-1  from: 1  to: 8535

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         |||||||||      :::||||||:::|||||||||||||:::||||||
    1001 ATGGGACAAGATAATTCTACCCCTATCTCCCTCACTCTAAATCACTGGAG 1050
```

```
  17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
1051 AGATGTGAGAACAAGGGCTCACAATCTATCCGTGGAAATCAAAAAGGGAA 1100

  34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
1101 AATGGCAGACTTTCTGTTCCTCCGAGTGGCCCACATTCGGCGTGGGGTGG 1150

  51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
1151 CCACCGGAGGGAACTTTTAATCTCTCTGTCATTTTTGCAGTTAAAAAGAT 1200

  67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
1201 TGTCTTTCAGGAGAAC...GGGGGACATCCGGACCAAGTTCCATATATCG 1247

  84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
1248 TGGTATGGCAGGACCTCGCCCAGAATCCCCCACCATGGGTG..... 1288

 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
1289 .............CCAGCCTCCGCCAAGGTCGCTGTTGTCTCTGATAC 1323

 117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
1324 CCGAAGACCAGTTGCGGGGAGG......CCATCAGCTCCTCCCCGACCCC 1367

 134 rgAspArgGlyAlaAlaAsp.........LeuAlaGlyThrProThr... 146
1368 CCATCTACCCGGCAACAGACGACTTACTCCTCCTCTCTGAACCCACGCCC 1417

 147 ..............CysSerProThrProLeuSerSerThrGlyCysCy 158
1418 CCGCCCTATCCGGCGGCACTGCCACCCCCTCTGGCCCCTCAGGCGATCGG 1467

 158 sGluGlyThrSerAlaProProGlyAlaProValValGluGlyProAlaA 175
1468 ACCGCCGTCAGGCCAGATGCCCGATAGTAGCGATCCTGAGGGGCCAGCCG 1517

 175 laGlyThrArgSerArgArgGlyAlaThrPro.........GluArgThr 188
1518 CTGGGACCAGGAGTCGCCGTGCCCGCAGTCCAGCAGACAACTCGGGTCCT 1567

 189 AspGluIleAlaIleLeuProLeuArgThrTyrGlyProProMetProGl 205
1568 GACTCCACTGTGATTTTGCCCCTCCGAGCCATAGGACCCCCGGCCGAGCC 1617

 205 yGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAspLeuT 222
1618 CAATGGCCTGGTCCCTCTACAATATTGGCCTTTTTCCTCAGCAGATCTTT 1667

 222 yrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeu 238
1668 ATAATTGGAAATCTAATCATCCCTCTTTTTCTGAAAACCCAGCAGGTCTC 1717

 239 ThrGlyLeuValGluSerLeuMetPheSerHisGlnProThrThrTrpAspAs 255
1718 ACGGGGCTCCTTGAGTCTCTTATGTTCTCCCATCAGCCCACTTGGGACGA 1767

 255 pCysGlnGlnLeuLeuGlnThrLeuPheThrThrGluGluArgGluArgI 272
1768 TTGCCAACAGCTCCTACAGATTCTTTTCACCACTGAGGAACGGGAAAGAA 1817

 272 leLeuLeuGluAlaArgLysAsnValProGlyAlaAspGlyArgProThr 288
1818 TTCTCCTGGAGGCCCGCAAAAATGTCCTTGGGGACAATGGGGCCCCTACA 1867

 289 GlnLeuGlnAsnGluIleAspMetGlyPheProLeuThrArgProGlyTr 305
1868 CAGCTCGAGAACCTCATTAATGAGGCCTTCCCCCTCAATCGACCTCACTG 1917

 305 pAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArgGlnA 322
```

```
1918 GGATTACAACACAGCCGCAGGTAGGGAGCGTCTTCTGGTCTACCGCCGGA 1967

 322 laLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAla 338
1968 CTCTAGTGGCAGGTCTCAAAGGGGCAGCTCGGCGTCCTACCAATTTGGCT 2017

 339 LysValArgGluValMetGlnGlyProAsnGluProProSerValPheLe 355
2018 AAGGTAAGAGAGGTCTTGCAGGGACCGGCAGAACCCCCTTCGGTTTTCTT 2067

 355 uGluArgLeuMetGluAlaPheArgArgPheThrProPheAspProThrS 372
2068 AGAACGCCTGATGGAGGCCTATAGGAGATACACTCCGTTTGATCCCTCTT 2117

 372 erGluAlaGlnLysAlaSerValAlaLeuAlaPheIleGlyGlnSerAla 388
2118 CTGAGGGACAACAGGCTGCGGTCGCCATGGCCTTTATCGGACAGTCAGCC 2167

 389 LeuAspIleArgLysLysLeuGlnArgLeuGluGlyLeuGlnGluAlaGl 405
2168 CCAGATATCAAGAAAAAGTTACAGAGGCTAGAGGGGCTCCAGGACTATTC 2217

 405 uLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArgGluT 422
2218 CTTACAAGATTTAGTAAAAGAGGCAGAAAAGGTGTACCATAAGAGAGAGA 2267

 422 hrGluGluGluLysGluGlnArgLysGluLysGluArgGluGluArgGlu 438
2268 CAGAAGAAGAAAGACAAGAAAGAGAAAAAAAGGAGGCAGAAGAAAAGGAG 2317

 439 GluArgArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAl 455
2318 AGGCGGCGCGATAGGCCGAAGAAAAAAAACTTGACTAAAATTCTGGCCGC 2367

 455 aValValGluGlyLysSerSerArgGluArgGluArgAspPheArgLysI 472
2368 AGTAGTA.............AGTAGAGAA..................... 2383

 472 leArgSerGlyProArgGlnSerGlyAsnLeuGlyAsn............ 484
2384 ..GGGTCCACAGGTAGGCAGACAGGGAACCTGAGCAACCAGGCAAAGAAG 2431

 485 ..............ArgThrProLeuAspLysAspGlnCysAlaTyrCy 496
2432 ACACCTAGGGATGGAAGACCTCCACTAGACAAAGACCAGTGCGCATACTG 2481

 496 sLysGluLysGlyHisTrpAlaArgAsnCysProLysLysGlyAsn...L 512
2482 TAAAGAGAAGGGCCATTGGGCAAGAGAATGTCCCCGAAAAAAACACGTCA 2531

 512 ysGlyProLysValLeuAlaLeuGlu 520
2532 GAGAAGCCAAGGTTCTAGCCCTAGAT 2557
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-3

seq_documentation_block:
; Sequence 3, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23

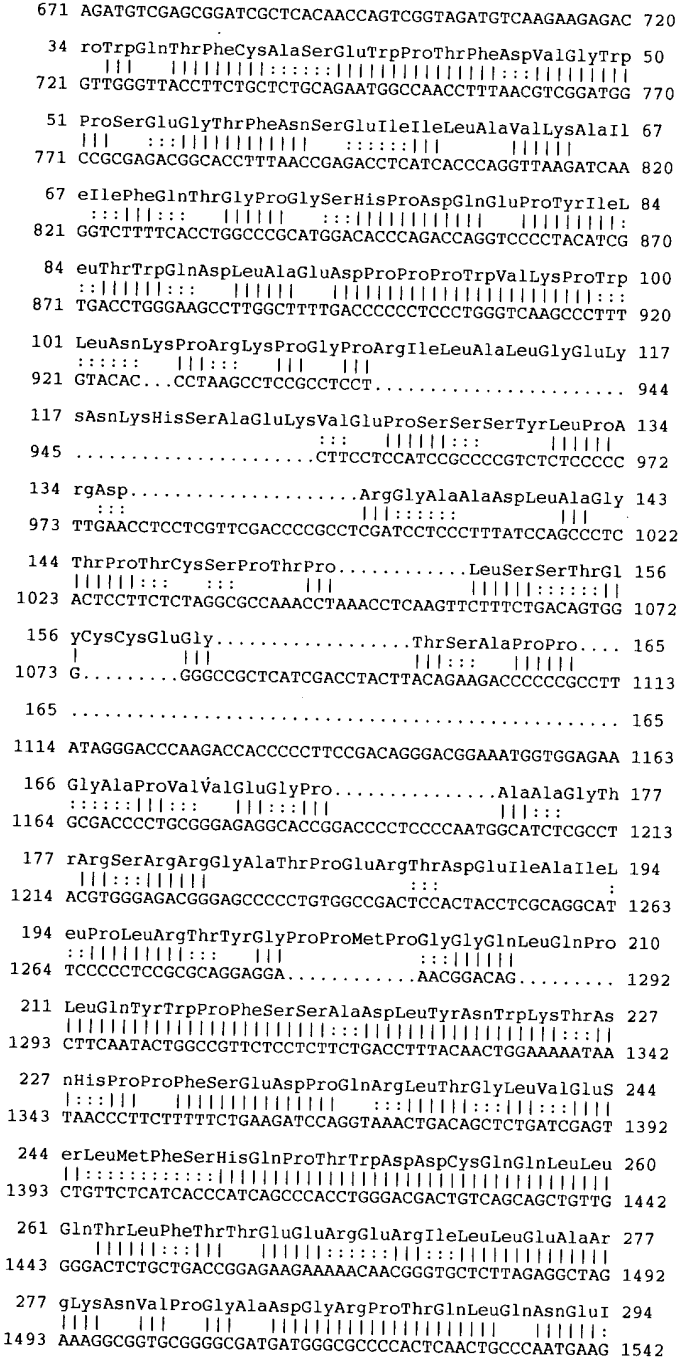
```
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 7308
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
US-09-011-745-3


alignment_scores:
            Quality: 1355.50              Length:    578
              Ratio:  3.440               Gaps:      15
 Percent Similarity:  68.166   Percent Identity: 48.962

alignment_block:
 US-09-171-553B-4 x US-09-011-745-3   ..

 Align seg 1/1  to: US-09-011-745-3  from: 1  to: 7308

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         |||||||||||||||||||||||||||||||||||||||||||   ||||||
    1567 ATGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAA 1616

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
         :::|||:::      |||||||||   ||||||:::|||||||||
    1617 AGATGTCGAGCGGATCGCTCACAACCAGTCGGTAGATGTCAAGAAGAGAC 1666

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
           |||   ||||||||:::::::|||||||||||||||:::|||||||||
    1667 GTTGGGTTACCTTCTGCTCTGCAGAATGGCCAACCTTTAACGTCGGATGG 1716

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||   :::||||||||||||| ~ ::::::|||      ||||||
    1717 CCGCGAGACGGCACCTTTAACCGAGACCTCATCACCCAGGTTAAGATCAA 1766

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
         :::|||:::   ||||||   :::||||||||||||   |||||||||:
    1767 GGTCTTTTCACCTGGCCCGCATGGACACCCAGACCAGGTCCCCTACATCG 1816

      84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
         ::||||||:::   ||||||   |||||||||||||||||||||||:::
    1817 TGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGCCCTTT 1866

     101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
         ::::::   |||:::   |||   |||
    1867 GTACAC...CCTAAGCCTCCGCCTCCT..................... 1890

     117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
                                 :::   ||||||:::      ||||||
    1891 .....................CTTCCTCCATCCGCCCCGTCTCTCCCCC 1918

     134 rgAsp.....................ArgGlyAlaAlaAspLeuAlaGly 143
           :::                     |||:::::          |||
    1919 TTGAACCTCCTCGTTCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTC 1968

     144 ThrProThrCysSerProThrPro.............LeuSerSerThrGl 156
         |||||||:::   :::      |||           ||||||:::::||
    1969 ACTCCTTCTCTAGGCGCCAAACCTAAACCTCAAGTTCTTTCTGACAGTGG 2018

     156 yCysCysGluGly..................ThrSerAlaProPro.... 165
         |          |||                 |||:::   ||||||
    2019 G.........GGGCCGCTCATCGACCTACTTACAGAAGACCCCCCGCCTT 2059

     165 ..................................................... 165

    2060 ATAGGGACCCAAGACCACCCCCTTCCGACAGGGACGGAAATGGTGGAGAA 2109

     166 GlyAlaProValValGluGlyPro..............AlaAlaGlyTh 177
         ::::::|||:::   |||:::|||               |||:::
    2110 GCGACCCCTGCGGGAGAGGCACCGGACCCCTCCCCAATGGCATCTCGCCT 2159

     177 rArgSerArgArgGlyAlaThrProGluArgThrAspGluIleAlaIleL 194
          |||:::||||||                  :::                 :
    2160 ACGTGGGAGACGGGAGCCCCCTGTGGCCGACTCCACTACCTCGCAGGCAT 2209

     194 euProLeuArgThrTyrGlyProProMetProGlyGlyGlnLeuGlnPro 210
         ::|||||||||:::   |||            :::||||||
    2210 TCCCCCTCCGCGCAGGAGGA.............AACGGACAG......... 2238

     211 LeuGlnTyrTrpProPheSerSerAlaAspLeuTyrAsnTrpLysThrAs 227
         |||||||||||||||||||||||:::|||||||||||||||||||:::|||
    2239 CTTCAATACTGGCCGTTCTCCTCTTCTGACCTTTACAACTGGAAAAATAA 2288

     227 nHisProProPheSerGluAspProGlnArgLeuThrGlyLeuValGluS 244
         |:::|||   |||||||||||||||   :::||||||:::|||:::||||
    2289 TAACCCTTCTTTTTCTGAAGATCCAGGTAAACTGACAGCTCTGATCGAGT 2338

     244 erLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
         ||:::::::::::|||||||||||||||||||||||||||||||||||||
    2339 CTGTTCTCATCACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTG 2388

     261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaAr 277
           ||||||:::|||   ||||||:::::|||:::||||||||||||||
    2389 GGGACTCTGCTGACCGGAGAAGAAAAACAACGGGTGCTCTTAGAGGCTAG 2438

     277 gLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluI 294
         ||||  |||  |||  ||||||||||||||||||||   ||||||:
    2439 AAAGGCGGTGCGGGGCGATGATGGGCGCCCCACTCAACTGCCCAATGAAG 2488

     294 leAspMetGlyPheProLeuThrArgProGlyTrpAspTyrAsnThrAla 310
         ::|||   :::|||||||||   ||||||   ||||||||||:::|||
    2489 TCGATGCCGCTTTTCCCCTCGAGCGCCCAGACTGGGATTACACCACCCAG 2538

     311 GluGlyArgGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLe 327
         ||||||:::   |||      |||||||||   |||:::||||||||
    2539 GCAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCGGGTCT 2588

     327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluValM 344
         |::::::|||:::|||   ||||||||||||||||||||:::   :::
    2589 CCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGAATAA 2638

     344 etGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
           |||||||||||||||   ||||||:::||||||||||||||||   |||
    2639 CACAAGGGCCCAATGAGTCTCCCTCGGCCTTCCTAGAGAGACTTAAGGAA 2688

     361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAl 377
         |||:::|||||:::||||||:::|||||||   :::   :::|||:::::
    2689 GCCTATCGCAGGTACACTCCTTATGACCCTGAGGACCCAGGGCAAGAAAC 2738

     377 aSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysL 394
         ::::|||:::::::::||||||   |||||||||   ||||||   :::|
    2739 TAATGTGTCTATGTCTTTCATTTGGCAGTCTGCCCCAGACATTGGGAGAA 2788

     394 ysLeuGlnArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuVal 410
         |||||:::|||||||||   |||::::::      |||   |||||||||
    2789 AGTTAGAGAGGTTAGAAGATTTAAAAAACAAGACGCTTGGAGATTTGGTT 2838

     411 ArgGluAlaGluLysValTyrTyrArgArgGluThrGluGluGluLysGl 427
         ||||||||||||||||:::::   :::|||||||||   ||||||:::||
    2839 AGAGAGGCAGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAAAGAGA 2888

     427 uGlnArgLysGluLysGluArgGluGluArgGluGluArg..... 440
         |:::|||   ::::::|||   ||||||:::|||||||||
    2889 GGAACGTATCAGGAGAGAAACAGAGGAAAAAGAAGAACGCCGTAGGACAG 2938

     441 .....................ArgAspArgArgGlnGluLysAsnLeu 449
                               ||||||||||||:::::::::::::::
    2939 AGGATGAGCAGAAAGAGAAAGAAAGAGATCGTAGGAGACATAGAGAGATG 2988

     450 ThrLysIleLeuAlaAlaValValGluGlyLysSerSerArgGluArgGl 466
         :::|||:::||||||:::||||||:::|||:::::::::
    2989 AGCAAGCTATTGGCCACTGTCGTTAGTGGACAGAAACAG........... 3027
```

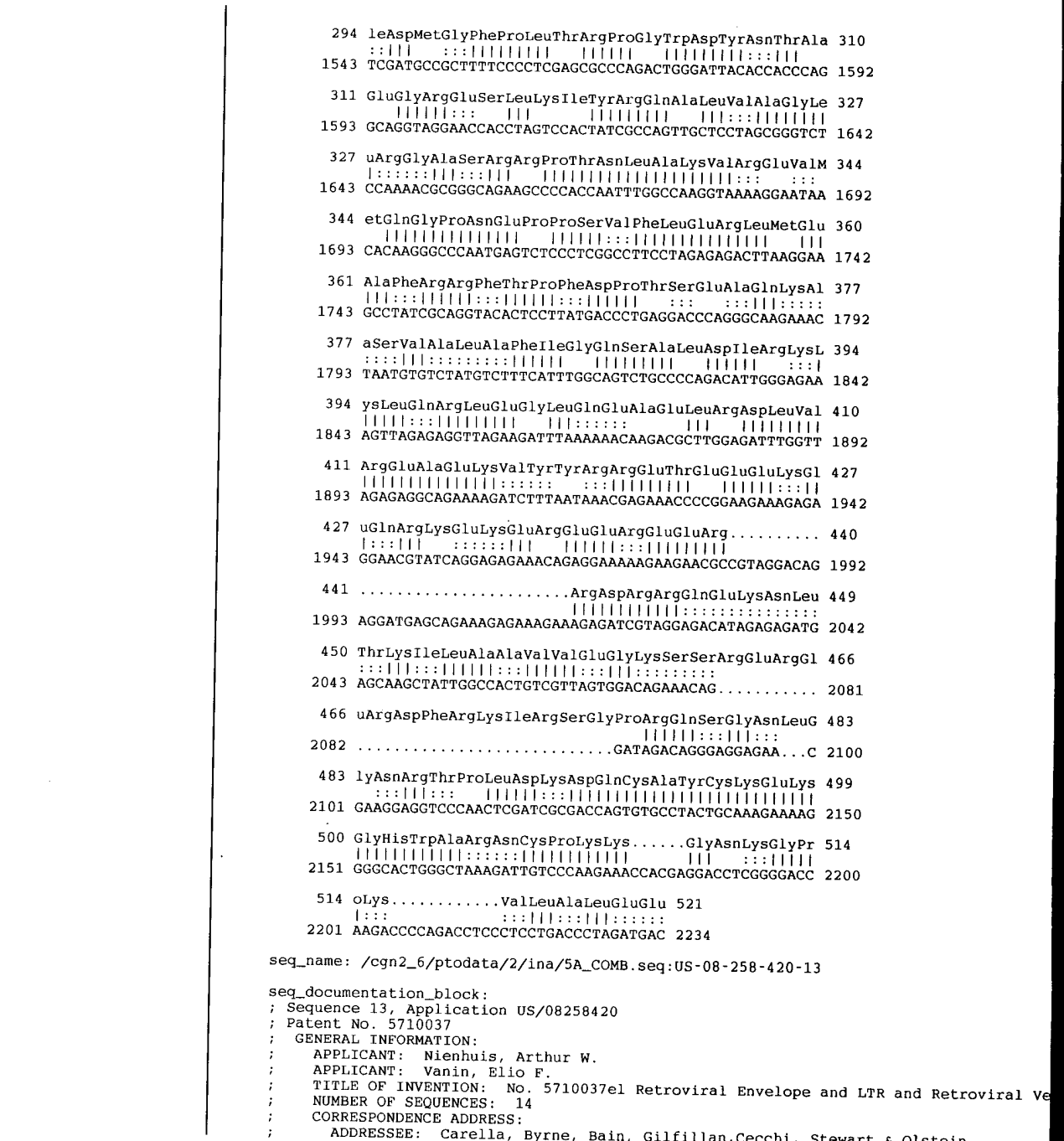
```
    466 uArgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuG 483
                                    ||||||:::|||:::
   3028 ..........................GATAGACAGGGAGGAGAA...C 3046

    483 lyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLys 499
          :::|||:::   ||||||:::|||||||||||||||||||||||||||
   3047 GAAGGAGGTCCCAACTCGATCGCGACCAGTGTGCCTACTGCAAAGAAAAG 3096

    500 GlyHisTrpAlaArgAsnCysProLysLys......GlyAsnLysGlyPr 514
        ||||||||||||:::::||||||||||||      |||   ::::||||
   3097 GGGCACTGGGCTAAAGATTGTCCCAAGAAACCACGAGGACCTCGGGGACC 3146

    514 oLys............ValLeuAlaLeuGluGlu 521
        |:::           :::|||:::|||::::::
   3147 AAGACCCCAGACCTCCCTCCTGACCCTAGATGAC 3180

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-4

seq_documentation_block:
; Sequence 4, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 7308
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
US-09-011-745-4


alignment_scores:
            Quality: 1355.50          Length:    578
              Ratio:  3.440            Gaps:     15
 Percent Similarity:  68.166   Percent Identity: 48.962

alignment_block:
 US-09-171-553B-4 x US-09-011-745-4   ..

 Align seg 1/1  to: US-09-011-745-4  from: 1  to: 7308

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         ||||||||||||||||||||||||||||||||||||||||   ||||||
    1567 ATGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAA 1616

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
         :::|||:::     |||||||||   |||||||:||||||||||
    1617 AGATGTCGAGCGGATCGCTCACAACCAGTCGGTAGATGTCAAGAAGAGAC 1666

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
           |||   |||||||||:::::::||||||||||||||:::|||||||||
    1667 GTTGGGTTACCTTCTGCTCTGCAGAATGGCCAACCTTTAACGTCGGATGG 1716

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||   :::|||||||||||||   :::::::|||      ||||||
    1717 CCGCGAGACGGCACCTTTAACCGAGACCTCATCACCCAGGTTAAGATCAA 1766

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84

     :::|||:::   ||||||   :::|||||||||||||   |||||||||:
1767 GGTCTTTTCACCTGGCCCGCATGGACACCCAGACCAGGTCCCCTACATCG 1816

  84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
     ::|||||||:::   ||||||   ||||||||||||||||||||||||:::
1817 TGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGCCCTTT 1866

 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
     ::::::  |||:::   |||   |||
1867 GTACAC...CCTAAGCCTCCGCCTCCT..................... 1890

 117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
                              :::   |||||||:::      ||||||
1891 .....................CTTCCTCCATCCGCCCCGTCTCTCCCCC 1918

 134 rgAsp.....................ArgGlyAlaAlaAspLeuAlaGly 143
       :::                    |||::::::          |||
1919 TTGAACCTCCTCGTTCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTC 1968

 144 ThrProThrCysSerProThrPro.............LeuSerSerThrGl 156
     |||||||:::   :::      |||             ||||||::::::||
1969 ACTCCTTCTCTAGGCGCCAAACCTAAACCTCAAGTTCTTTCTGACAGTGG 2018

 156 yCysCysGluGly..................ThrSerAlaProPro.... 165
     |    |||                      |||:::   ||||||
2019 G.........GGGCCGCTCATCGACCTACTTACAGAAGACCCCCCGCCTT 2059

 165 ..................................................... 165

2060 ATAGGGACCCAAGACCACCCCCTTCCGACAGGGACGGAAATGGTGGAGAA 2109

 166 GlyAlaProValValGluGlyPro..............AlaAlaGlyTh 177
     ::::::|||:::   |||:::|||               |||:::
2110 GCGACCCCTGCGGGAGAGGCACCGGACCCCTCCCCAATGGCATCTCGCCT 2159

 177 rArgSerArgArgGlyAlaThrProGluArgThrAspGluIleAlaIleL 194
      |||:::||||||          :::              :
2160 ACGTGGGAGACGGGAGCCCCCTGTGGCCGACTCCACTACCTCGCAGGCAT 2209

 194 euProLeuArgThrTyrGlyProProMetProGlyGlyGlnLeuGlnPro 210
     ::|||||||||:::   |||            :::||||||
2210 TCCCCCTCCGCGCAGGAGGA............AACGGACAG......... 2238

 211 LeuGlnTyrTrpProPheSerSerAlaAspLeuTyrAsnTrpLysThrAs 227
     ||||||||||||||||||||||||:::|||||||||||||||||||:::||
2239 CTTCAATACTGGCCGTTCTCCTCTTCTGACCTTTACAACTGGAAAAATAA 2288

 227 nHisProProPheSerGluAspProGlnArgLeuThrGlyLeuValGluS 244
     |:::|||   ||||||||||||||   :::||||||:::|||:::||||
2289 TAACCCTTCTTTTTCTGAAGATCCAGGTAAACTGACAGCTCTGATCGAGT 2338

 244 erLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
     ||:::::::::::::|||||||||||||||||||||||||||||||||||
2339 CTGTTCTCATCACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTG 2388

 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaAr 277
     |||||||:::|||   |||||||:::::|||:::||||||||||||||||
2389 GGGACTCTGCTGACCGGAGAAGAAAAACAACGGGTGCTCTTAGAGGCTAG 2438

 277 gLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluI 294
     ||||   |||   |||   ||||||||||||||||||||   ||||||:
2439 AAAGGCGGTGCGGGGCGATGATGGGCGCCCCACTCAACTGCCCAATGAAG 2488

 294 leAspMetGlyPheProLeuThrArgProGlyTrpAspTyrAsnThrAla 310
     ::|||   :::||||||||||   ||||||   |||||||||:::|||
2489 TCGATGCCGCTTTTCCCCTCGAGCGCCCAGACTGGGATTACACCACCCAG 2538

 311 GluGlyArgGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLe 327
     ||||||:::   |||      |||||||||   |||:::||||||||
2539 GCAGGACGCAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCGGGTCT 2588

 327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluValM 344
     |:::::::|||:::|||   |||||||||||||||||||||:::   :::
```

```
 2589 CCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGAATAA 2638

  344 etGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
      ||||||||||||||||   ||||||:::|||||||||||||||   |||
 2639 CACAAGGGCCCAATGAGTCTCCCTCGGCCTTCCTAGAGAGACTTAAGGAA 2688

  361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAl 377
      |||:::||||||:::||||||:::||||||   :::   :::|||:::::
 2689 GCCTATCGCAGGTACACTCCTTATGACCCTGAGGACCCAGGGCAAGAAAC 2738

  377 aSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysL 394
      ::::|||:::::::::||||||   |||||||||   ||||||   :::|
 2739 TAATGTGTCTATGTCTTTCATTTGGCAGTCTGCCCCAGACATTGGGAGAA 2788

  394 ysLeuGlnArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuVal 410
      |||||:::||||||||||   |||::::::      |||   |||||||||
 2789 AGTTAGAGAGGTTAGAAGATTTAAAAAACAAGACGCTTGGAGATTTGGTT 2838

  411 ArgGluAlaGluLysValTyrTyrArgArgGluThrGluGluGluLysGl 427
      ||||||||||||||:::::   :::||||||||||   ||||||:::|||
 2839 AGAGAGGCAGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAAAGAGA 2888

  427 uGlnArgLysGluLysGluArgGluGluArgGluGluArg..... 440
      |:::|||   :::::|||   ||||||:::||||||||
 2889 GGAACGTATCAGGAGAGAAACAGAGGAAAAAGAAGAACGCCGTAGGACAG 2938

  441 .....ArgAspArgArgGlnGluLysAsnLeu 449
                ||||||||||||:::::::::::::::
 2939 AGGATGAGCAGAAAGAGAAAGAAAGAGATCGTAGGAGACATAGAGAGATG 2988

  450 ThrLysIleLeuAlaAlaValValGluGlyLysSerSerArgGluArgGl 466
      :::|||:::|||||||:::|||||||:::|||::::::::::
 2989 AGCAAGCTATTGGCCACTGTCGTTAGTGGACAGAAACAG..... 3027

  466 uArgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuG 483
                          |||||||:::|||:::
 3028 .....GATAGACAGGGAGGAGAA...C 3046

  483 lyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLys 499
        :::|||:::   ||||||:::|||||||||||||||||||||||||||||
 3047 GAAGGAGGTCCCAACTCGATCGCGACCAGTGTGCCTACTGCAAAGAAAAG 3096

  500 GlyHisTrpAlaArgAsnCysProLysLys......GlyAsnLysGlyPr 514
      ||||||||||||:::::|||||||||||||      |||   :::||||||
 3097 GGGCACTGGGCTAAAGATTGTCCCAAGAAACCACGAGGACCTCGGGGACC 3146

  514 oLys.....ValLeuAlaLeuGluGlu 521
      |:::          :::|||:::|||::::::
 3147 AAGACCCCAGACCTCCCTCCTGACCCTAGATGAC 3180

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-2

seq_documentation_block:
; Sequence 2, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 7616
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
US-09-011-745-2


alignment_scores:
            Quality: 1355.50                Length:    578
              Ratio:  3.440                   Gaps:     15
 Percent Similarity:  68.166   Percent Identity: 48.962

alignment_block:
 US-09-171-553B-4 x US-09-011-745-2   ..

 Align seg 1/1  to: US-09-011-745-2  from: 1  to: 7616

    1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
      ||||||||||||||||||||||||||||||||||||||||||||   ||||||
  728 ATGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAA 777

   17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
      :::|||:::      |||||||||   |||||||:::|||||||||
  778 AGATGTCGAGCGGATCGCTCACAACCAGTCGGTAGATGTCAAGAAGAGAC 827

   34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
        |||   |||||||||:::::|||||||||||||||||:::|||||||||
  828 GTTGGGTTACCTTCTGCTCTGCAGAATGGCCAACCTTTAACGTCGGATGG 877

   51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
      |||   :::||||||||||||   ::::::|||      ||||||
  878 CCGCGAGACGGCACCTTTAACCGAGACCTCATCACCCAGGTTAAGATCAA 927

   67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
      :::|||:::   ||||||   :::||||||||||||   |||||||||:
  928 GGTCTTTTCACCTGGCCCGCATGGACACCCAGACCAGGTCCCCTACATCG 977

   84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
      ::|||||:::   ||||||   |||||||||||||||||||||||||:::
  978 TGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGCCCTTT 1027

  101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
      ::::::   |||:::   |||   |||
 1028 GTACAC...CCTAAGCCTCCGCCTCCT..... 1051

  117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
                              :::   ||||||:::      ||||||
 1052 .....CTTCCTCCATCCGCCCCGTCTCTCCCCC 1079

  134 rgAsp.....ArgGlyAlaAlaAspLeuAlaGly 143
      :::          |||::::::         |||
 1080 TTGAACCTCCTCGTTCGACCCCGCCTCGATCCTCCCTTTATCCAGCCCTC 1129

  144 ThrProThrCysSerProThrPro.....LeuSerSerThrGl 156
      |||||||:::   :::      |||           ||||||::::::||
 1130 ACTCCTTCTCTAGGCGCCAAACCTAAACCTCAAGTTCTTTCTGACAGTGG 1179

  156 yCysCysGluGly.....ThrSerAlaProPro.... 165
      |        |||           |||:::   ||||||
 1180 G.....GGGCCGCTCATCGACCTACTTACAGAAGACCCCCCGCCTT 1220

  165 ..... 165

 1221 ATAGGGACCCAAGACCACCCCCTTCCGACAGGGACGGAAATGGTGGAGAA 1270

  166 GlyAlaProValValGluGlyPro.....AlaAlaGlyTh 177
      ::::::|||:::   |||:::|||           |||:::
 1271 GCGACCCCTGCGGGAGAGGCACCGGACCCCTCCCCAATGGCATCTCGCCT 1320

  177 rArgSerArgArgGlyAlaThrProGluArgThrAspGluIleAlaIleL 194
      |||:::||||||                :::              :
 1321 ACGTGGGAGACGGGAGCCCCCTGTGGCCGACTCCACTACCTCGCAGGCAT 1370
```

```
 194 euProLeuArgThrTyrGlyProProMetProGlyGlyGlnLeuGlnPro 210
     ::|||||||||:::   |||            :::|||||||
1371 TCCCCCTCCGCGCAGGAGGA.............AACGGACAG......... 1399

 211 LeuGlnTyrTrpProPheSerSerAlaAspLeuTyrAsnTrpLysThrAs 227
     ||||||||||||||||||||||:::||||||||||||||||||:::||
1400 CTTCAATACTGGCCGTTCTCCTCTTCTGACCTTTACAACTGGAAAAATAA 1449

 227 nHisProProPheSerGluAspProGlnArgLeuThrGlyLeuValGluS 244
     |:::|||   |||||||||||||||   :::|||||||:::|||:::||||
1450 TAACCCTTCTTTTTCTGAAGATCCAGGTAAACTGACAGCTCTGATCGAGT 1499

 244 erLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
     ||:::::::::::|||||||||||||||||||||||||||||||||||||
1500 CTGTTCTCATCACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTG 1549

 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaAr 277
     ||||||:::|||   ||||||:::::|||:::|||||||||||||||
1550 GGGACTCTGCTGACCGGAGAAGAAAAACAACGGGTGCTCTTAGAGGCTAG 1599

 277 gLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluI 294
     ||||   |||   |||   ||||||||||||||||||||   ||||||:
1600 AAAGGCGGTGCGGGGCGATGATGGGCGCCCCACTCAACTGCCCAATGAAG 1649

 294 leAspMetGlyPheProLeuThrArgProGlyTrpAspTyrAsnThrAla 310
     ::|||   :::|||||||||   ||||||   ||||||||||:::|||
1650 TCGATGCCGCTTTTCCCCTCGAGCGCCCAGACTGGGATTACACCACCCAG 1699

 311 GluGlyArgGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLe 327
     |||||||::   |||      |||||||||   |||:::|||||||||
1700 GCAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCGGGTCT 1749

 327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluValM 344
     |:::::::|||:::|||   |||||||||||||||||||||:::   :::
1750 CCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGAATAA 1799

 344 etGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
       ||||||||||||||||   ||||||:::|||||||||||||||   |||
1800 CACAAGGGCCCAATGAGTCTCCCTCGGCCTTCCTAGAGAGACTTAAGGAA 1849

 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAl 377
     |||:::|||||||:::|||||||:::||||||   :::   :::|||:::::
1850 GCCTATCGCAGGTACACTCCTTATGACCCTGAGGACCCAGGGCAAGAAAC 1899

 377 aSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysL 394
     ::::|||:::::::::::||||||   |||||||||   ||||||   :::|
1900 TAATGTGTCTATGTCTTTCATTTGGCAGTCTGCCCCAGACATTGGGAGAA 1949

 394 ysLeuGlnArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuVal 410
     ||||||:::|||||||||   |||:::::::      |||   |||||||||
1950 AGTTAGAGAGGTTAGAAGATTTAAAAAACAAGACGCTTGGAGATTTGGTT 1999

 411 ArgGluAlaGluLysValTyrTyrArgArgGluThrGluGluGluLysGl 427
     ||||||||||||||||::::::   :::||||||||||   ||||||:::||
2000 AGAGAGGCAGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAAAGAGA 2049

 427 uGlnArgLysGluLysGluArgGluGluArgGluGluArg..... 440
     |:::|||   :::::::|||   ||||||:::|||||||||
2050 GGAACGTATCAGGAGAGAAACAGAGGAAAAAGAAGAACGCCGTAGGACAG 2099

 441 .....ArgAspArgArgGlnGluLysAsnLeu 449
                              ||||||||||||:::::::::::::::
2100 AGGATGAGCAGAAAGAGAAAGAAAGAGATCGTAGGAGACATAGAGAGATG 2149

 450 ThrLysIleLeuAlaAlaValValGluGlyLysSerSerArgGluArgGl 466
     :::|||:::||||||:::||||||:::|||:::::::::
2150 AGCAAGCTATTGGCCACTGTCGTTAGTGGACAGAAACAG.......... 2188

 466 uArgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuG 483
                                  ||||||:::|||:::
2189 .........................GATAGACAGGGAGGAGAA...C 2207

 483 lyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLys 499
        :::|||:::   ||||||:::||||||||||||||||||||||||||||
2208 GAAGGAGGTCCCAACTCGATCGCGACCAGTGTGCCTACTGCAAAGAAAAG 2257

 500 GlyHisTrpAlaArgAsnCysProLysLys......GlyAsnLysGlyPr 514
     ||||||||||||:::::||||||||||||      |||   :::||||||
2258 GGGCACTGGGCTAAAGATTGTCCCAAGAAACCACGAGGACCTCGGGGACC 2307

 514 oLys.............ValLeuAlaLeuGluGlu 521
     |:::             :::|||:::|||::::::
2308 AAGACCCCAGACCTCCCTCCTGACCCTAGATGAC 2341
```

```
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-850-961-1

seq_documentation_block:
; Sequence 1, Application US/08850961
; Patent No. 6013517
;  GENERAL INFORMATION:
;     APPLICANT:  Respess, James G.
;     APPLICANT:  De Polo, Nicholas J.
;     APPLICANT:  Chada, Sunil
;     APPLICANT:  Sauter, Sybille
;     APPLICANT:  Bodner, Mordechai
;     APPLICANT:  Driver, David A.
;     TITLE OF INVENTION:  CROSSLESS RETROVIRAL VECTORS
;     NUMBER OF SEQUENCES:  45
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Chiron Corporation, Intellectual Property - R440
;       STREET:  P.O. Box 8097
;       CITY:  Emeryville
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94662-8097
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/850,961
;       FILING DATE:  05-MAY-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kruse, No. 6013517man J.
;       REGISTRATION NUMBER:  35,235
;       REFERENCE/DOCKET NUMBER:  930049.424C4 / 1147.005
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (510) 601-3520
;       TELEFAX:  (510) 655-3542
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8332 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-850-961-1


alignment_scores:
            Quality: 1355.50              Length:    578
              Ratio:  3.440                Gaps:     15
 Percent Similarity:  68.166   Percent Identity: 48.962

alignment_block:
 US-09-171-553B-4 x US-08-850-961-1   ..

 Align seq 1/1  to: US-08-850-961-1  from: 1  to: 8332

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         ||||||||||||||||||||||||||||||||||||||||   ||||||
     621 ATGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAA 670

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
          :::|||:::      |||||||||   ||||||:::|||||||||
```



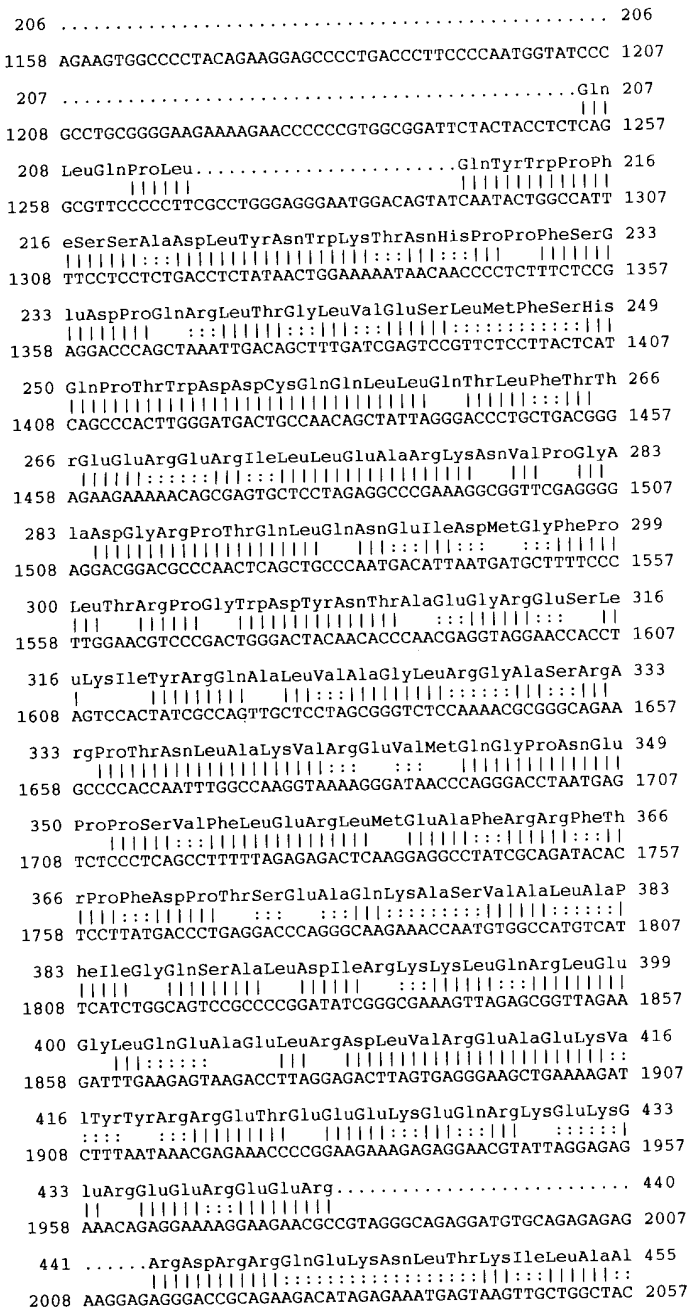


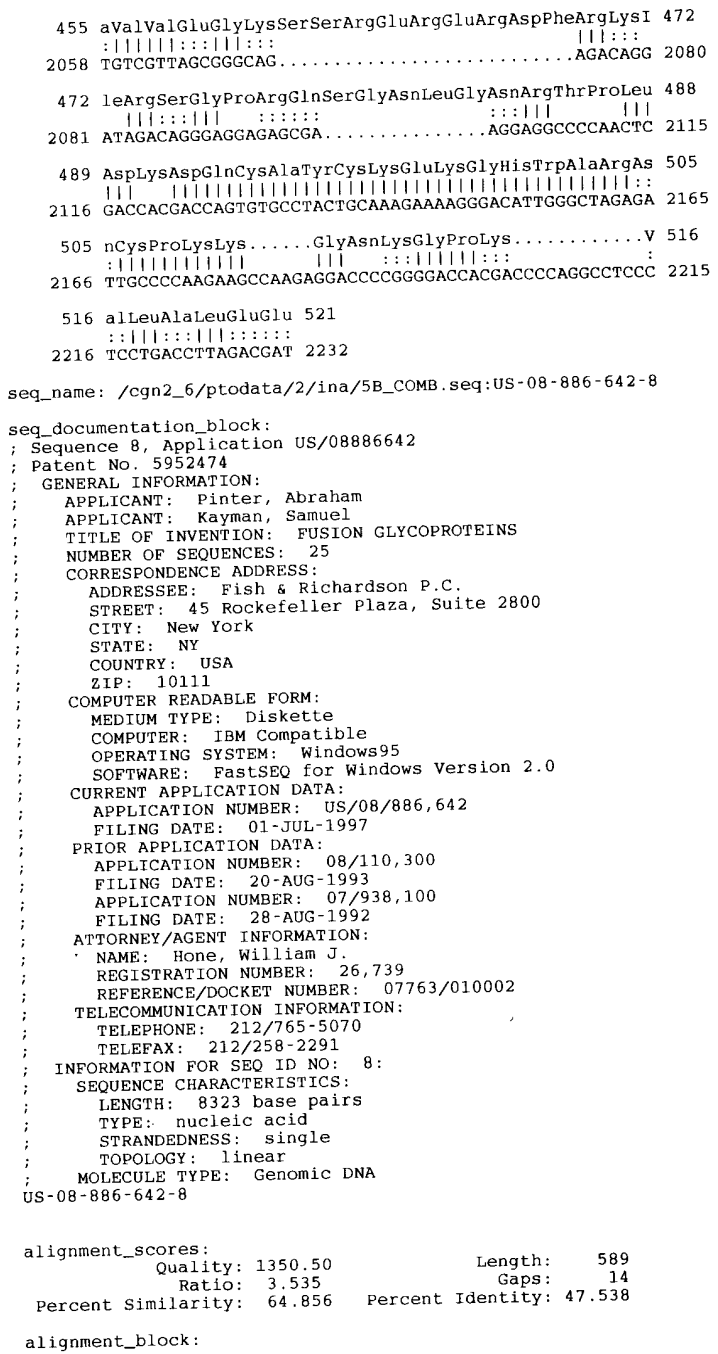
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-13

seq_documentation_block:
; Sequence 13, Application US/08258420
; Patent No. 5710037
; GENERAL INFORMATION:
; APPLICANT: Nienhuis, Arthur W.
; APPLICANT: Vanin, Elio F.
; TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Ve
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,Cecchi, Stewart & Olstein

```
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch diskette
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,420
;      FILING DATE:  10-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Olstein, Elliot M.
;      REGISTRATION NUMBER:  24,025
;      REFERENCE/DOCKET NUMBER:  271010-208
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8202 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  viral genome
US-08-258-420-13


alignment_scores:
            Quality: 1354.50              Length:    578
              Ratio:  3.438                Gaps:     15
 Percent Similarity:  68.166   Percent Identity: 48.789

alignment_block:
 US-09-171-553B-4 x US-08-258-420-13/rev   ..

 Align seg 1/1  to reverse of: US-08-258-420-13  from: 1  to: 8202

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         ||||||||||||||||||||||||||||||||||||||||   ||||||
    7146 ATGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAA 7097

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
          :::|||:::      |||||||||   ||||||:::|||||||||
    7096 AGATGTCGAGCGGATCGCTCACAACCAGTCGGTAGATGTCAAGAAGAGAC 7047

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
           |||   |||||||||:::::|||||||||||||||||:::::||||||
    7046 GTTGGGTTACCTTCTGCTCTGCAGAATGGCCAACCTTTAACATCGGATGG 6997

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||   :::|||||||||||||   ::::::|||      ||||||
    6996 CCGCGAGACGGCACCTTTAACCGAGACCTCATCACCCAGGTTAAGATCAA 6947

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
          :::|||:::   ||||||   :::|||||||||||||   |||||||||:
    6946 GGTCTTTTCACCTGGCCCGCATGGACACCCAGACCAGGTCCCCTACATCG 6897

      84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
         ::|||||:::   ||||||   ||||||||||||||||||||||||:::
    6896 TGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGCCCTTT 6847

     101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
         ::::::   |||:::   |||   |||
    6846 GTACAC...CCTAAGCCTCCGCCTCCT..................... 6823

     117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
                                 :::   ||||||:::      ||||||
    6822 .....................CTTCCTCCATCCGCCCCGTCTCTCCCCC 6795

     134 rgAsp.....................ArgGlyAlaAlaAspLeuAlaGly 143
          :::                    |||::::::          |||
    6794 TTGAACCTCCTCGTTCGACCCCGCCTCGCTCCTCCCTTTATCCAGCCCTC 6745

     144 ThrProThrCysSerProThrPro............LeuSerSerThrGl 156
         |||||||:::   :::      |||           ||||||:::::||
    6744 ACTCCTTCTCTAGGCGCCAAACCTAAACCTCAAGTTCTTTCTGACAGTGG 6695

     156 yCysCysGluGly.................ThrSerAlaProPro.... 165
         |         |||                 |||:::   ||||||
    6694 G.........GGGCCGCTCATCGACCTACTTACAGAAGACCCCCCGCCTT 6654

     165 ..............................................  165

    6653 ATAGGGACCCAAGACCACCCCCTTCCGACAGGGACGGAAATGGTGGAGAA 6604

     166 GlyAlaProValValGluGlyPro..............AlaAlaGlyTh 177
         :::::::|||:::   |||:::|||              |||:::
    6603 GCGACCCCTGCGGGAGAGGCACCGGACCCCTCCCCAATGGCATCTCGCCT 6554

     177 rArgSerArgArgGlyAlaThrProGluArgThrAspGluIleAlaIleL 194
          |||:::||||||                 :::               :
    6553 ACGTGGGAGACGGGAGCCCCCTGTGGCCGACTCCACTACCTCGCAGGCAT 6504

     194 euProLeuArgThrTyrGlyProProMetProGlyGlyGlnLeuGlnPro 210
         ::|||||||||:::  |||           :::||||||
    6503 TCCCCCTCCGCGCAGGAGGA............AACGGACAG......... 6475

     211 LeuGlnTyrTrpProPheSerSerAlaAspLeuTyrAsnTrpLysThrAs 227
         ||||||||||||||||||||||||:::|||||||||||||||||||:::||
    6474 CTTCAATACTGGCCGTTCTCCTCTTCTGACCTTTACAACTGGAAAAATAA 6425

     227 nHisProProPheSerGluAspProGlnArgLeuThrGlyLeuValGluS 244
         |:::|||   ||||||||||||||   :::||||||:::|||:::||||
    6424 TAACCCTTCTTTTTCTGAAGATCCAGGTAAACTGACAGCTCTGATCGAGT 6375

     244 erLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
         ||:::::::::::::|||||||||||||||||||||||||||||||||||
    6374 CTGTTCTCATCACCCATCAGCCCACCTGGGACGACTGTCAGCAGCTGTTG 6325

     261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaAr 277
         ||||||:::|||   ||||||:::::|||:::|||||||||||||||
    6324 GGGACTCTGCTGACCGGAGAAGAAAAACAACGGGTGCTCTTAGAGGCTAG 6275

     277 gLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluI 294
         ||||   |||   |||   ||||||||||||||||||||   ||||||:
    6274 AAAGGCGGTGCGGGGCGATGATGGGCGCCCCACTCAACTGCCCAATGAAG 6225

     294 leAspMetGlyPheProLeuThrArgProGlyTrpAspTyrAsnThrAla 310
         ::|||   :::|||||||||   ||||||   |||||||||:::|||
    6224 TCGATGCCGCTTTTCCCCTCGAGCGCCCAGACTGGGATTACACCACCCAG 6175

     311 GluGlyArgGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLe 327
         ||||||||:::   |||      |||||||||   |||:::|||||||||
    6174 GCAGGTAGGAACCACCTAGTCCACTATCGCCAGTTGCTCCTAGCGGGTCT 6125

     327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluValM 344
         |:::::::|||:::|||   |||||||||||||||||||||:::   :::
    6124 CCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGAATAA 6075

     344 etGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
         |||||||||||||||   ||||||:::|||||||||||||||   |||
    6074 CACAAGGGCCCAATGAGTCTCCCTCGGCCTTCCTAGAGAGACTTAAGGAA 6025

     361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAl 377
         |||:::||||||:::||||||:::||||||   :::   :::|||:::::
    6024 GCCTATCGCAGGTACACTCCTTATGACCCTGAGGACCCAGGGCAAGAAAC 5975
```

```
     377 aSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysL 394
         ::::|||::::::::::|||||   |||||||||   ||||||   :::|
    5974 TAATGTGTCTATGTCTTTCATTTGGCAGTCTGCCCCAGACATTGGGAGAA 5925

     394 ysLeuGlnArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuVal 410
         |||||:::|||||||||   |||::::::      |||   |||||||||
    5924 AGTTAGAGAGGTTAGAAGATTTAAAAAACAAGACGCTTGGAGATTTGGTT 5875

     411 ArgGluAlaGluLysValTyrTyrArgArgGluThrGluGluGluLysGl 427
         ||||||||||||||||:::::   :::|||||||||   ||||||:::||
    5874 AGAGAGGCAGAAAAGATCTTTAATAAACGAGAAACCCCGGAAGAAAGAGA 5825

     427 uGlnArgLysGluLysGluArgGluGluArgGluGluArg........... 440
         |:::|||   :::::::|||   ||||||:::|||||||||
    5824 GGAACGTATCAGGAGAGAAACAGAGGAAAAAGAAGAACGCCGTAGGACAG 5775

     441 ......................ArgAspArgArgGlnGluLysAsnLeu 449
                                |||||||||||||:::::::::::::::
    5774 AGGATGAGCAGAAAGAGAAAGAAAGAGATCGTAGGAGACATAGAGAGATG 5725

     450 ThrLysIleLeuAlaAlaValValGluGlyLysSerSerArgGluArgGl 466
         :::|||:::|||||||:::|||||||:::|||:::::::::
    5724 AGCAAGCTATTGGCCACTGTCGTTAGTGGACAGAAACAG........... 5686

     466 uArgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuG 483
                                       ||||||:::|||:::
    5685 ............................GATAGACAGGGAGGAGAA...C 5667

     483 lyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLys 499
           :::|||:::   ||||||:::|||||||||||||||||||||||||||||
    5666 GAAGGAGGTCCCAACTCGATCGCGACCAGTGTGCCTACTGCAAAGAAAAG 5617

     500 GlyHisTrpAlaArgAsnCysProLysLys......GlyAsnLysGlyPr 514
         ||||||||||||:::::::|||||||||||||      |||   :::|||||
    5616 GGGCACTGGGCTAAAGATTGTCCCAAGAAACCACGAGGACCTCGGGGACC 5567

     514 oLys............ValLeuAlaLeuGluGlu 521
         |:::            :::|||:::|||::::::
    5566 AAGACCCCAGACCTCCCTCCTGACCCTAGATGAC 5533

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-8

seq_documentation_block:
; Sequence 8, Application US/08110300A
; Patent No. 5643756
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,300A
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461

;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-110-300A-8


alignment_scores:
            Quality: 1350.50              Length:    589
              Ratio:  3.535                Gaps:     14
 Percent Similarity:  64.856   Percent Identity: 47.538

alignment_block:
 US-09-171-553B-4 x US-08-110-300A-8   ..

 Align seg 1/1  to: US-08-110-300A-8  from: 1  to: 8323

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         |||||||||:::|||||||||||||||||||||||||||||||||||||
     619 ATGGGCCAGGCTGTTACCACCCCCTTAAGTTTGACTTTAGACCACTGGAA 668

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
          :::|||:::      ||||||||||||||||||:::|||:::|||
     669 GGATGTCGAACGGACAGCCCACAACCTGTCGGTAGAGGTTAGAAAAAGGC 718

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
           |||   |||||||||:::::|||||||||||||||:::|||||||||
     719 GCTGGGTTACATTCTGCTCTGCAGAATGGCCAACCTTCAACGTCGGATGG 768

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||   :::|||||||||||||   :::||||||      ||||||
     769 CCACGAGACGGCACTTTTAACCCAGACATTATTACACAGGTTAAGATCAA 818

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
         :::|||:::   ||||||   :::|||||||||||   |||||||||:
     819 GGTCTTCTCACCTGGCCCACATGGACATCCGGATCAGGTCCCCTACATCG 868

      84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
         ::||||||:::   :::|||   ||||||||||||||||||:::|||:::
     869 TGACCTGGGAAGCTATAGCAGTAGACCCCCCTCCCTGGGTCAGACCCTTC 918

     101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
         ::::::   |||:::   |||            |||:::|||
     919 GTGCAC...CCTAAACCTCCC............CTCTCTCTT........ 945

     117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
                                        ||||||:::      ||||||
     946 ........................CCCCCTTCAGCCCCCTCTCTCCCAC 970

     134 rgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
           :::                          |||      |||
     971 CTGAA.........................CCCCCACTCTCGACCCCG 993

     151 ProLeuSerSerThrGlyCysCysGluGlyThrSerAlaProProGlyAl 167
         |||   ||||||
     994 CCCCAGTCCTCCCTC..............................TA 1010

     167 aProValValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAlaT 184
          |||::::::   :::|||      :::|||:::   |||      :::
    1011 TCCGGCTCTCACTTCTCCT...TTAAACACCAAACCTAGGCCTCAAGTCC 1057

     184 hrProGluArgThrAspGluIleAlaIleLeu..................... 194
           |||:::            :::      |||
    1058 TTCCTGATAGCGGAGGACCACTCATTGATCTACTCACGGAGGACCCTCCG 1107

     195 ProLeuArgThrTyrGlyProProMetProGlyGly.............. 206
         |||   |||      |||||||||   |||   |||
    1108 CCTTACCGGGACCCAGGGCCACCCTCTCCTGACGGGAACGGCGATAGCGG 1157
```

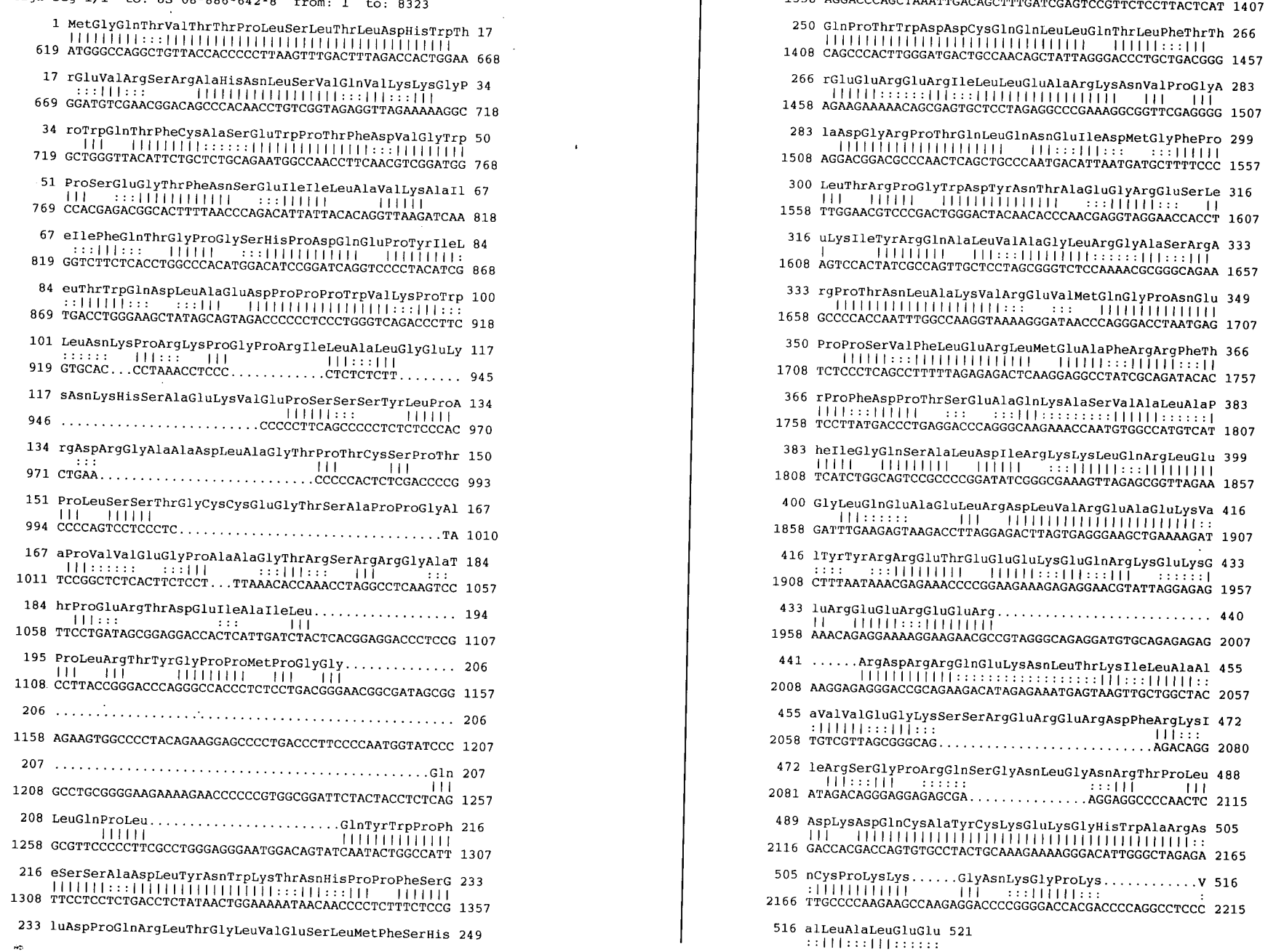




```
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-8

seq_documentation_block:
; Sequence 8, Application US/08886642
; Patent No. 5952474
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  45 Rockefeller Plaza, Suite 2800
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/886,642
;      FILING DATE:  01-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/110,300
;      FILING DATE:  20-AUG-1993
;      APPLICATION NUMBER:  07/938,100
;      FILING DATE:  28-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hone, William J.
;      REGISTRATION NUMBER:  26,739
;      REFERENCE/DOCKET NUMBER:  07763/010002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212/765-5070
;      TELEFAX:  212/258-2291
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-886-642-8


alignment_scores:
                Quality: 1350.50              Length:    589
                  Ratio:  3.535                Gaps:     14
 Percent Similarity:  64.856   Percent Identity: 47.538

alignment_block:
```

US-09-171-553B-4 x US-08-886-642-8 ..

Align seq 1/1 to: US-08-886-642-8 from: 1 to: 8323



```
    2216 TCCTGACCTTAGACGAT 2232

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-8

seq_documentation_block:
; Sequence 8, Application PC/TUS9308041
;  GENERAL INFORMATION:
;    APPLICANT:  The Public Health Research Institute of the City of
;    APPLICANT:  New York, Inc.
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08041
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/938,100
;      FILING DATE:  28-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/_____
;      FILING DATE:  20-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-08041-8


alignment_scores:
            Quality: 1350.50              Length:    589
              Ratio:  3.535                Gaps:     14
 Percent Similarity:  64.856   Percent Identity: 47.538

alignment_block:
 US-09-171-553B-4 x PCT-US93-08041-8   ..

 Align seg 1/1  to: PCT-US93-08041-8  from: 1  to: 8323

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         |||||||||:::||||||||||||||||||||||||||||||||||||||
     619 ATGGGCCAGGCTGTTACCACCCCCTTAAGTTTGACTTTAGACCACTGGAA 668

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
          :::|||:::      |||||||||||||||||||:::|||:::|||
     669 GGATGTCGAACGGACAGCCCACAACCTGTCGGTAGAGGTTAGAAAAAGGC 718

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
           |||   |||||||||:::::|||||||||||||||||:::|||||||||
     719 GCTGGGTTACATTCTGCTCTGCAGAATGGCCAACCTTCAACGTCGGATGG 768

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||   :::|||||||||||||   :::||||||      ||||||
     769 CCACGAGACGGCACTTTTAACCCAGACATTATTACACAGGTTAAGATCAA 818

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
          :::|||:::   ||||||   :::|||||||||||||   |||||||||:
     819 GGTCTTCTCACCTGGCCCACATGGACATCCGGATCAGGTCCCCTACATCG 868

      84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
         ::|||||:::   :::|||   |||||||||||||||||:::|||:::
     869 TGACCTGGGAAGCTATAGCAGTAGACCCCCCTCCCTGGGTCAGACCCTTC 918

      101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
          ::::::   |||:::  |||           |||:::|||
     919 GTGCAC...CCTAAACCTCCC.....CTCTCTCTT..... 945

      117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
                                      ||||||:::     ||||||
     946 .....CCCCCTTCAGCCCCCTCTCTCCCAC 970

      134 rgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
          :::                          |||      |||
     971 CTGAA.....CCCCCACTCTCGACCCCG 993

      151 ProLeuSerSerThrGlyCysCysGluGlyThrSerAlaProProGlyAl 167
          |||   ||||||
     994 CCCCAGTCCTCCCTC.....TA 1010

      167 aProValValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAlaT 184
          |||::::::   :::|||      :::|||:::   |||     ':::
    1011 TCCGGCTCTCACTTCTCCT...TTAAACACCAAACCTAGGCCTCAAGTCC 1057

      184 hrProGluArgThrAspGluIleAlaIleLeu..... 194
          |||:::           :::      |||
    1058 TTCCTGATAGCGGAGGACCACTCATTGATCTACTCACGGAGGACCCTCCG 1107

      195 ProLeuArgThrTyrGlyProProMetProGlyGly..... 206
          |||   |||      |||||||||   |||   |||
    1108 CCTTACCGGGACCCAGGGCCACCCTCTCCTGACGGGAACGGCGATAGCGG 1157

      206 ..... 206

    1158 AGAAGTGGCCCCTACAGAAGGAGCCCCTGACCCTTCCCCAATGGTATCCC 1207

      207 .....Gln 207
                                                     |||
    1208 GCCTGCGGGGAAGAAAAGAACCCCCCGTGGCGGATTCTACTACCTCTCAG 1257

      208 LeuGlnProLeu.....GlnTyrTrpProPh 216
          ||||||                ||||||||||||||
    1258 GCGTTCCCCCTTCGCCTGGGAGGGAATGGACAGTATCAATACTGGCCATT 1307

      216 eSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerG 233
          |||||||:::|||||||||||||||||||:::|||:::|||   |||||||
    1308 TTCCTCCTCTGACCTCTATAACTGGAAAAATAACAACCCCTCTTTCTCCG 1357

      233 luAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHis 249
          ||||||||   :::|||||||:::|||:::||||||:::::::::::|||
    1358 AGGACCCAGCTAAATTGACAGCTTTGATCGAGTCCGTTCTCCTTACTCAT 1407

      250 GlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrTh 266
          |||||||||||||||||||||||||||||||   ||||||:::|||
    1408 CAGCCCACTTGGGATGACTGCCAACAGCTATTAGGGACCCTGCTGACGGG 1457

      266 rGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyA 283
          ||||||:::::|||:::|||||||||||||||||   |||   |||
    1458 AGAAGAAAAACAGCGAGTGCTCCTAGAGGCCCGAAAGGCGGTTCGAGGGG 1507

      283 laAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPhePro 299
          ||||||||||||||||||   |||:::|||:::   :::||||||
    1508 AGGACGGACGCCCAACTCAGCTGCCCAATGACATTAATGATGCTTTTCCC 1557
```

```
     300 LeuThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLe 316
         |||   ||||||   ||||||||||||||   :::||||||:::   ||
    1558 TTGGAACGTCCCGACTGGGACTACAACACCCAACGAGGTAGGAACCACCT 1607

     316 uLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgA 333
         |      |||||||||   |||:::|||||||||::::::|||:::|||
    1608 AGTCCACTATCGCCAGTTGCTCCTAGCGGGTCTCCAAAACGCGGGCAGAA 1657

     333 rgProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGlu 349
           ||||||||||||||||||||:::   :::   ||||||||||||||||
    1658 GCCCCACCAATTTGGCCAAGGTAAAAGGGATAACCCAGGGACCTAATGAG 1707

     350 ProProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheTh 366
         ||||||:::|||||||||||||||   ||||||:::||||||:::||
    1708 TCTCCCTCAGCCTTTTTAGAGAGACTCAAGGAGGCCTATCGCAGATACAC 1757

     366 rProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaP 383
         ||||:::||||||   :::   :::|||:::::::::|||||:::::|
    1758 TCCTTATGACCCTGAGGACCCAGGGCAAGAAACCAATGTGGCCATGTCAT 1807

     383 heIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 399
         |||||   |||||||||   ||||||   :::||||||:::|||||||||
    1808 TCATCTGGCAGTCCGCCCCGGATATCGGGCGAAAGTTAGAGCGGTTAGAA 1857

     400 GlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysVa 416
           |||::::::      |||   |||||||||||||||||||||||::
    1858 GATTTGAAGAGTAAGACCTTAGGAGACTTAGTGAGGGAAGCTGAAAAGAT 1907

     416 lTyrTyrArgArgGluThrGluGluGluLysGluGlnArgLysGluLysG 433
         ::::   :::|||||||||   ||||||:::|||:::|||   ::::::|
    1908 CTTTAATAAACGAGAAACCCCGGAAGAAAGAGAGGAACGTATTAGGAGAG 1957

     433 luArgGluGluArgGluGluArg..........................  440
         ||   ||||||:::|||||||||
    1958 AAACAGAGGAAAAGGAAGAACGCCGTAGGGCAGAGGATGTGCAGAGAGAG 2007

     441 ......ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAl 455
               ||||||||||||:::::::::::::::::::|||:::||||||::
    2008 AAGGAGAGGGACCGCAGAAGACATAGAGAAATGAGTAAGTTGCTGGCTAC 2057

     455 aValValGluGlyLysSerSerArgGluArgGluArgAspPheArgLysI 472
         :||||||:::|||:::                          |||:::
    2058 TGTCGTTAGCGGGCAG.........................AGACAGG 2080

     472 leArgSerGlyProArgGlnSerGlyAsnLeuGlyAsnArgThrProLeu 488
           |||:::|||   ::::::                ::::||      |||
    2081 ATAGACAGGGAGGAGAGCGA..............AGGAGGCCCCAACTC 2115

     489 AspLysAspGlnCysAlaTyrCysLysGluLysGlyHisTrpAlaArgAs 505
         |||   ||||||||||||||||||||||||||||||||||||||||||::
    2116 GACCACGACCAGTGTGCCTACTGCAAAGAAAAGGGACATTGGGCTAGAGA 2165

     505 nCysProLysLys......GlyAsnLysGlyProLys...........V 516
         :||||||||||||      |||   :::||||||:::           :
    2166 TTGCCCCAAGAAGCCAAGAGGACCCCGGGGACCACGACCCCAGGCCTCCC 2215

     516 alLeuAlaLeuGluGlu 521
         ::|||:::|||::::::
    2216 TCCTGACCTTAGACGAT 2232

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-9

seq_documentation_block:
; Sequence 9, Application US/08110300A
; Patent No. 5643756
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,300A
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-110-300A-9


alignment_scores:
           Quality: 1350.50              Length:    589
             Ratio:  3.535                Gaps:     14
 Percent Similarity:  64.856   Percent Identity: 47.538

alignment_block:
 US-09-171-553B-4 x US-08-110-300A-9/rev   ..

 Align seg 1/1  to reverse of: US-08-110-300A-9  from: 1  to: 10367

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         |||||||||:::|||||||||||||||||||||||||||||||||||||
    6977 ATGGGCCAGGCTGTTACCACCCCCTTAAGTTTGACTTTAGACCACTGGAA 6928

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
          :::|||:::      ||||||||||||||||||:::|||:::|||
    6927 GGATGTCGAACGGACAGCCCACAACCTGTCGGTAGAGGTTAGAAAAAGGC 6878

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
           |||   |||||||||:::::|||||||||||||||||:::|||||||||
    6877 GCTGGGTTACATTCTGCTCTGCAGAATGGCCAACCTTCAACGTCGGATGG 6828

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||   :::|||||||||||||   :::|||||||      ||||||
    6827 CCACGAGACGGCACTTTTAACCCAGACATTATTACACAGGTTAAGATCAA 6778

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
          :::|||:::   ||||||   :::|||||||||||||   |||||||||:
    6777 GGTCTTCTCACCTGGCCCACATGGACATCCGGATCAGGTCCCCTACATCG 6728

      84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
         ::||||||:::   :::|||   ||||||||||||||||||:::|||:::
    6727 TGACCTGGGAAGCTATAGCAGTAGACCCCCCTCCCTGGGTCAGACCCTTC 6678

     101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
         ::::::   |||:::   |||            |||:::|||
    6677 GTGCAC...CCTAAACCTCCC............CTCTCTCTT........ 6651

     117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
                                        ||||||:::      ||||||
    6650 ........................CCCCCTTCAGCCCCCTCTCTCCCAC 6626
```

```
 134 rgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
6625 CTGAA.........................CCCCCACTCTCGACCCCG 6603

 151 ProLeuSerSerThrGlyCysCysGluGlyThrSerAlaProProGlyAl 167
6602 CCCCAGTCCTCCCTC.............................TA 6586

 167 aProValValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAlaT 184
6585 TCCGGCTCTCACTTCTCCT...TTAAACACCAAACCTAGGCCTCAAGTCC 6539

 184 hrProGluArgThrAspGluIleAlaIleLeu..................... 194
6538 TTCCTGATAGCGGAGGACCACTCATTGATCTACTCACGGAGGACCCTCCG 6489

 195 ProLeuArgThrTyrGlyProProMetProGlyGly.............. 206
6488 CCTTACCGGGACCCAGGGCCACCCTCTCCTGACGGGAACGGCGATAGCGG 6439

 206 ..................................................... 206
6438 AGAAGTGGCCCCTACAGAAGGAGCCCCTGACCCTTCCCCAATGGTATCCC 6389

 207 ...................................................Gln 207
6388 GCCTGCGGGGAAGAAAAGAACCCCCCGTGGCGGATTCTACTACCTCTCAG 6339

 208 LeuGlnProLeu.......................GlnTyrTrpProPh 216
6338 GCGTTCCCCCTTCGCCTGGGAGGGAATGGACAGTATCAATACTGGCCATT 6289

 216 eSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerG 233
6288 TTCCTCCTCTGACCTCTATAACTGGAAAAATAACAACCCCTCTTTCTCCG 6239

 233 luAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHis 249
6238 AGGACCCAGCTAAATTGACAGCTTTGATCGAGTCCGTTCTCCTTACTCAT 6189

 250 GlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrTh 266
6188 CAGCCCACTTGGGATGACTGCCAACAGCTATTAGGGACCCTGCTGACGGG 6139

 266 rGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyA 283
6138 AGAAGAAAAACAGCGAGTGCTCCTAGAGGCCCGAAAGGCGGTTCGAGGGG 6089

 283 laAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPhePro 299
6088 AGGACGGACGCCCAACTCAGCTGCCCAATGACATTAATGATGCTTTTCCC 6039

 300 LeuThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLe 316
6038 TTGGAACGTCCCGACTGGGACTACAACACCCAACGAGGTAGGAACCACCT 5989

 316 uLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgA 333
5988 AGTCCACTATCGCCAGTTGCTCCTAGCGGGTCTCCAAAACGCGGGCAGAA 5939

 333 rgProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGlu 349
5938 GCCCCACCAATTTGGCCAAGGTAAAAGGGATAACCCAGGGACCTAATGAG 5889

 350 ProProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheTh 366
5888 TCTCCCTCAGCCTTTTTAGAGAGACTCAAGGAGGCCTATCGCAGATACAC 5839

 366 rProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaP 383
5838 TCCTTATGACCCTGAGGACCCAGGGCAAGAAACCAATGTGGCCATGTCAT 5789

 383 heIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 399
5788 TCATCTGGCAGTCCGCCCCGGATATCGGGCGAAAGTTAGAGCGGTTAGAA 5739

 400 GlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysVa 416
5738 GATTTGAAGAGTAAGACCTTAGGAGACTTAGTGAGGGAAGCTGAAAAGAT 5689

 416 lTyrTyrArgArgGluThrGluGluGluLysGluGlnArgLysGluLysG 433
5688 CTTTAATAAACGAGAAACCCCGGAAGAAAGAGAGGAACGTATTAGGAGAG 5639

 433 luArgGluGluArgGluGluArg.............................. 440
5638 AAACAGAGGAAAAGGAAGAACGCCGTAGGGCAGAGGATGTGCAGAGAGAG 5589

 441 ......ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAl 455
5588 AAGGAGAGGGACCGCAGAAGACATAGAGAAATGAGTAAGTTGCTGGCTAC 5539

 455 aValValGluGlyLysSerSerArgGluArgGluArgAspPheArgLysI 472
5538 TGTCGTTAGCGGGCAG.............................AGACAGG 5516

 472 leArgSerGlyProArgGlnSerGlyAsnLeuGlyAsnArgThrProLeu 488
5515 ATAGACAGGGAGGAGAGCGA...............AGGAGGCCCCAACTC 5481

 489 AspLysAspGlnCysAlaTyrCysLysGluLysGlyHisTrpAlaArgAs 505
5480 GACCACGACCAGTGTGCCTACTGCAAAGAAAAGGGACATTGGGCTAGAGA 5431

 505 nCysProLysLys......GlyAsnLysGlyProLys............V 516
5430 TTGCCCCAAGAAGCCAAGAGGACCCCGGGGACCACGACCCCAGGCCTCCC 5381

 516 alLeuAlaLeuGluGlu 521
5380 TCCTGACCTTAGACGAT 5364
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-9

```
seq_documentation_block:
; Sequence 9, Application US/08886642
; Patent No. 5952474
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  45 Rockefeller Plaza, Suite 2800
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/886,642
;      FILING DATE:  01-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/110,300
;      FILING DATE:  20-AUG-1993
;      APPLICATION NUMBER:  07/938,100
;      FILING DATE:  28-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hone, William J.
;      REGISTRATION NUMBER:  26,739
```

```
;       REFERENCE/DOCKET NUMBER:  07763/010002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212/765-5070
;       TELEFAX:  212/258-2291
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10367 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
US-08-886-642-9


alignment_scores:
           Quality: 1350.50              Length:    589
             Ratio:  3.535                Gaps:     14
 Percent Similarity:  64.856   Percent Identity: 47.538

alignment_block:
 US-09-171-553B-4 x US-08-886-642-9/rev   ..

 Align seg 1/1  to reverse of: US-08-886-642-9  from: 1  to: 10367

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         ||||||||||:::||||||||||||||||||||||||||||||||||||
    6977 ATGGGCCAGGCTGTTACCACCCCCTTAAGTTTGACTTTAGACCACTGGAA 6928

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
         :::|||:::      ||||||||||||||||||:::|||:::|||
    6927 GGATGTCGAACGGACAGCCCACAACCTGTCGGTAGAGGTTAGAAAAAGGC 6878

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
         |||   |||||||||:::::::|||||||||||||||||:::|||||||||
    6877 GCTGGGTTACATTCTGCTCTGCAGAATGGCCAACCTTCAACGTCGGATGG 6828

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||   :::|||||||||||||   :::||||||      ||||||
    6827 CCACGAGACGGCACTTTTAACCCAGACATTATTACACAGGTTAAGATCAA 6778

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
         :::|||:::   ||||||   :::||||||||||   |||||||||:
    6777 GGTCTTCTCACCTGGCCCACATGGACATCCGGATCAGGTCCCCTACATCG 6728

      84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
         ::||||||:::   :::|||   ||||||||||||||||||:::|||:::
    6727 TGACCTGGGAAGCTATAGCAGTAGACCCCCCTCCCTGGGTCAGACCCTTC 6678

     101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLy 117
         ::::::   |||:::   |||            |||:::|||
    6677 GTGCAC...CCTAAACCTCCC...........CTCTCTCTT........ 6651

     117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
                                  ||||||:::      ||||||
    6650 .......................CCCCCTTCAGCCCCCTCTCTCCCAC 6626

     134 rgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
         :::                          |||      |||
    6625 CTGAA.......................CCCCCACTCTCGACCCCG 6603

     151 ProLeuSerSerThrGlyCysCysGluGlyThrSerAlaProProGlyAl 167
         |||   ||||||
    6602 CCCCAGTCCTCCCTC.............................TA 6586

     167 aProValValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAlaT 184
         |||:::::   :::|||      :::|||:::   |||      :::
    6585 TCCGGCTCTCACTTCTCCT...TTAAACACCAAACCTAGGCCTCAAGTCC 6539

     184 hrProGluArgThrAspGluIleAlaIleLeu................. 194
         |||:::           :::      |||
    6538 TTCCTGATAGCGGAGGACCACTCATTGATCTACTCACGGAGGACCCTCCG 6489

     195 ProLeuArgThrTyrGlyProProMetProGlyGly............. 206

         |||   |||      |||||||||   |||   |||
    6488 CCTTACCGGGACCCAGGGCCACCCTCTCCTGACGGGAACGGCGATAGCGG 6439

     206 ..................................................... 206

    6438 AGAAGTGGCCCCTACAGAAGGAGCCCCTGACCCTTCCCCAATGGTATCCC 6389

     207 .....................................................Gln 207
                                                             |||
    6388 GCCTGCGGGGAAGAAAAGAACCCCCCGTGGCGGATTCTACTACCTCTCAG 6339

     208 LeuGlnProLeu.....................GlnTyrTrpProPh 216
             ||||||                      ||||||||||||||
    6338 GCGTTCCCCCTTCGCCTGGGAGGGAATGGACAGTATCAATACTGGCCATT 6289

     216 eSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerG 233
         |||||||:::|||||||||||||||||||:::|||:::|||   |||||||
    6288 TTCCTCCTCTGACCTCTATAACTGGAAAAATAACAACCCCTCTTTCTCCG 6239

     233 luAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHis 249
         ||||||||   :::|||||||:::|||:::|||||:::::::::::::|||
    6238 AGGACCCAGCTAAATTGACAGCTTTGATCGAGTCCGTTCTCCTTACTCAT 6189

     250 GlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrTh 266
         ||||||||||||||||||||||||||||||||||   ||||||:::|||
    6188 CAGCCCACTTGGGATGACTGCCAACAGCTATTAGGGACCCTGCTGACGGG 6139

     266 rGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyA 283
         |||||:::::::|||:::|||||||||||||||||   |||   |||
    6138 AGAAGAAAAACAGCGAGTGCTCCTAGAGGCCCGAAAGGCGGTTCGAGGGG 6089

     283 laAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPhePro 299
         ||||||||||||||||||||   |||:::|||:::   :::||||||
    6088 AGGACGGACGCCCAACTCAGCTGCCCAATGACATTAATGATGCTTTTCCC 6039

     300 LeuThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLe 316
         |||   ||||||   ||||||||||||||   :::|||||:::   ||
    6038 TTGGAACGTCCCGACTGGGACTACAACACCCAACGAGGTAGGAACCACCT 5989

     316 uLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgA 333
         |      |||||||||   |||:::|||||||||:::::::|||:::|||
    5988 AGTCCACTATCGCCAGTTGCTCCTAGCGGGTCTCCAAAACGCGGGCAGAA 5939

     333 rgProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGlu 349
         ||||||||||||||||||||:::   :::   ||||||||||||||||
    5938 GCCCCACCAATTTGGCCAAGGTAAAAGGGATAACCCAGGGACCTAATGAG 5889

     350 ProProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheTh 366
         ||||||:::||||||||||||||||   ||||||:::||||||:::||
    5888 TCTCCCTCAGCCTTTTTAGAGAGACTCAAGGAGGCCTATCGCAGATACAC 5839

     366 rProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaP 383
         ||||:::||||||   :::   :::|||:::::::::|||||:::::::|
    5838 TCCTTATGACCCTGAGGACCCAGGGCAAGAAACCAATGTGGCCATGTCAT 5789

     383 heIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 399
         |||||   ||||||||   ||||||   :::||||||:::||||||||||
    5788 TCATCTGGCAGTCCGCCCCGGATATCGGGCGAAAGTTAGAGCGGTTAGAA 5739

     400 GlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysVa 416
         |||::::::      |||   ||||||||||||||||||||||||::
    5738 GATTTGAAGAGTAAGACCTTAGGAGACTTAGTGAGGGAAGCTGAAAAGAT 5689

     416 lTyrTyrArgArgGluThrGluGluGluLysGluGlnArgLysGluLysG 433
         ::::   :::||||||||||   ||||||:::|||:::|||   ::::::|
    5688 CTTTAATAAACGAGAAACCCCGGAAGAAAGAGAGGAACGTATTAGGAGAG 5639

     433 luArgGluGluArgGluGluArg.......................... 440
         ||   ||||||:::|||||||||
    5638 AAACAGAGGAAAAGGAAGAACGCCGTAGGGCAGAGGATGTGCAGAGAGAG 5589

     441 ......ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAl 455
               ||||||||||||:::::::::::::::::::|||:::|||||::
```

```
5588 AAGGAGAGGGACCGCAGAAGACATAGAGAAATGAGTAAGTTGCTGGCTAC 5539

 455 aValValGluGlyLysSerSerArgGluArgGluArgAspPheArgLysI 472
     :||||||:::|||:::                              |||:::
5538 TGTCGTTAGCGGGCAG.........................AGACAGG 5516

 472 leArgSerGlyProArgGlnSerGlyAsnLeuGlyAsnArgThrProLeu 488
       |||:::|||   ::::::               :::|||      |||
5515 ATAGACAGGGAGGAGAGCGA..............AGGAGGCCCCAACTC 5481

 489 AspLysAspGlnCysAlaTyrCysLysGluLysGlyHisTrpAlaArgAs 505
     |||   ||||||||||||||||||||||||||||||||||||||||||::
5480 GACCACGACCAGTGTGCCTACTGCAAAGAAAAGGGACATTGGGCTAGAGA 5431

 505 nCysProLysLys......GlyAsnLysGlyProLys...........V 516
     :||||||||||||      |||   :::|||||:::           :
5430 TTGCCCCAAGAAGCCAAGAGGACCCCGGGGACCACGACCCCAGGCCTCCC 5381

 516 alLeuAlaLeuGluGlu 521
     ::|||:::|||::::::
5380 TCCTGACCTTAGACGAT 5364
```

OM of: US-09-171-553B-5 to: Issued_Patents_NA:* out_format : pfs

Date: Feb 24, 2002 10:13 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09171553/runat_22022002_122826_11171/app_query.fasta_1.3312
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=feb22std.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09171553_@CGN1_1_158 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-171-553B-5
Query length: 1194
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 134.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-3	+	6099.50	11034.51	0.0	8132	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-1	+	5637.50	10193.87	0.0	8060	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-2	+	4489.50	8106.02	0.0	7333	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-075-272-1	+	4339.50	7830.73	0.0	8655	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-1	+	4319.00	7793.61	0.0	8535	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-13	-	3993.50	7201.81	0.0	8202	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-8	+	3980.00	7177.04	0.0	8323	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-8	+	3980.00	7177.04	0.0	8323	!
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-8	+	3980.00	7177.04	0.0	8323	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-9	-	3980.00	7173.98	0.0	10367	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-9	-	3980.00	7173.98	0.0	10367	!
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-9	-	3980.00	7173.98	0.0	10367	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-3	+	3967.00	7155.20	0.0	7308	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-4	+	3967.00	7155.20	0.0	7308	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-2	+	3967.00	7154.62	0.0	7616	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-850-961-1	+	3967.00	7153.37	0.0	8332	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-929-967-6	+	2510.00	4505.63	2.6e-243	6363	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-3	+	1782.50	3190.11	4.8e-170	3482	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-105-483-324	+	1554.50	2774.43	6.9e-147	3674	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-709-209-324	+	1554.50	2774.43	6.9e-147	3674	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-101-324	+	1554.50	2774.43	6.9e-147	3674	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-603-185-5	+	1510.50	2710.69	2.4e-143	1140	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-691-563C-57	+	1233.00	2195.35	1.2e-114	2391	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-991-789A-141	+	1145.00	2016.12	1.2e-104	9388	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-991-789A-227	+	1135.00	2011.12	2.3e-104	3646	
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US91-08254-1	+	923.00	1644.10	6.3e-84	949	
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US91-08254-2	-	923.00	1644.10	6.3e-84	949	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-7	+	827.00	1442.95	1.0e-72	6312	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-691-563C-53	+	814.50	1433.42	3.4e-72	2448	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-724-1	+	787.50	1394.73	4.9e-70	1158	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-969-1	+	787.50	1394.73	4.9e-70	1158	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-384-137-1	+	787.50	1394.73	4.9e-70	1158	!
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-470-006A-1	+	787.50	1394.73	4.9e-70	1158	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-691-563C-1	+	787.50	1394.73	4.9e-70	1158	!
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-200-990-1	+	787.50	1394.73	4.9e-70	1158	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-798-000-2	+	722.00	1274.61	2.4e-63	1237	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-691-563C-52	+	718.50	1259.06	1.8e-62	2389	!
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-798-000-3	+	712.50	1264.59	8.7e-63	735	! S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-798-000-1	+	709.50	1251.90	4.4e-62	1234	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-691-563C-89	+	650.50	1141.10	6.5e-56	1577	!
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-691-563C-61	+	619.00	1087.98	6.0e-53	11	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-121-321-1	+	608.50	1040.30	2.7e-50	9045	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-933-803A-1	+	608.50	1040.30	2.7e-50	904	
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-04910-8	+	604.00	1039.96	2.8e-50		
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-1	+	579.00	1004.77	2.6e-48	2462	

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-3

seq_documentation_block:
; Sequence 3, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-3

alignment_scores:
Quality: 6099.50 Length: 1191
Ratio: 5.173 Gaps: 1
Percent Similarity: 98.992 Percent Identity: 96.893

alignment_block:
US-09-171-553B-5 x US-08-766-528-3 ..

Align seg 1/1 to: US-08-766-528-3 from: 1 to: 8132

```
       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2160 GGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGT 2209

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerV 34
         ||||||||||||||||||||||||||||||||||||||:::|||||||
    2210 GGAGGGGCAACCAGTTGAGTTCCTGGTTGATACCGGAGCGAAACATTCAG 2259

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
         ||||||||||||||||||||||||||||||||:::|||||||||||||||
    2260 TGCTACTACAGCCATTAGGAAAACTAAAAGATAAAAAATCCTGGGTGATG 2309
```

```
  51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
     ||||||||||||||||:::|||||||||||||||||||||||||||||||
2310 GGTGCCACAGGGCAACAACAGTATCCATGGACTACCCGAAGAACAGTTGA 2359

  67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2360 CTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTCATACCTGAGTGCC 2409

  84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
     ||:::||||||||||||||||||||||||||||||||||||||||||||||
2410 CAGCACCCCTCTTAGGTAGAGACTTATTGACCAAGATGGGAGCACAAATT 2459

 101 SerPheGluGlnGlyArgProGluValSerValAsnAsnLysProIleTh 117
     ||||||||||||||:::|||||||||||||:::||||||||||||||||||
2460 TCTTTTGAACAAGGGAAACCAGAAGTGTCTGCAAATAACAAACCTATCAC 2509

 117 rValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnV 134
     ||||||||||||||||||||||||||||||||||||||||||||||||   |
2510 TGTGTTGACCCTCCAATTAGATGACGAATATCGACTATACTCTCCCCTAG 2559

 134 alLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGlnAla 150
     ||||||||||||||:::||||||   ||||||||||||||||||||||||
2560 TAAAGCCTGATCAAAATATACAATTCTGGTTGGAACAGTTTCCCCAAGCC 2609

 151 TrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnVa 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2610 TGGGCAGAAACCGCAGGGATGGGTTTGGCAAAGCAAGTTCCCCCACAAGT 2659

 167 lIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrProL 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2660 TATTCAACTGAAGGCCAGTGCCACACCAGTGTCAGTCAGACAGTACCCCT 2709

 184 euSerArgGluAlaArgGluGlyIleTrpProHisValGlnArgLeuIle 200
     |||||:::||||||:::||||||||||   ||||||||||||||||||||
2710 TGAGTAAAGAAGCTCAAGAAGGAATTCGGCCGCATGTCCAAAGATTAATC 2759

 201 GlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLe 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2760 CAACAGGGCATCCTAGTTCCTGTCCAATCTCCCTGGAATACTCCCCTGCT 2809

 217 uProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeuA 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2810 ACCGGTTAGAAAGCCTGGGACTAATGACTATCGACCAGTACAGGACTTGA 2859

 234 rgGluValAsnLysArgValGlnAspIleHisProThrValProAsnPro 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2860 GAGAGGTCAATAAACGGGTGCAGGATATACACCCAACAGTCCCGAACCCT 2909

 251 TyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLe 267
     ||||||||||||   ||||||||||||:::|||:::||||||||||||||
2910 TATAACCTCTTGTGTGCTCTCCCACCCCAACGGAGCTGGTATACAGTATT 2959

 267 uAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnP 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2960 GGACTTAAAGGATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAAC 3009

 284 roLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGln 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3010 CACTTTTTGCCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGGCAG 3059

 301 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePh 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3060 CTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTT 3109

 317 eAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3110 TGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATCCAACACCCTC 3159

 334 lnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAlaThr 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3160 AGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACC 3209

 351 LysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAs 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3210 AAACAGGACTGCTTAGAAGGCACGAAGGCACTACTGCTGGAATTGTCTGA 3259

 367 pLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluV 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3260 CCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGG 3309

 384 alThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrGlu 400
     ||||||||||||||||||||||||||   ||||||||||||||||||||||
3310 TAACATACTTGGGGTACAGTTTGCGGGACGGGCAGCGATGGCTGACGGAG 3359

 401 AlaArgLysLysThrValValGlnIleProAlaProThrThrAlaLysGl 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3360 GCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACA 3409

 417 nValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProG 434
     |:::||||||||||||||||||||||||||||||||||||||||||||||
3410 AATGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGG 3459

 434 lyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGly 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3460 GGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGG 3509

 451 GlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLy 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3510 GAATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATGCTATCAAAAA 3559

 467 sAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLysProP 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3560 GGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCT 3609

 484 heThrLeuTyrValAspGluArgLysGlyValAlaArgGlyValLeuThr 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3610 TTACCCTTTATGTGGATGAGCGTAAGGGAGTAGCCCGGGGAGTTTTAACC 3659

 501 GlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLe 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3660 CAAACCCTAGGACCATGGAGAAGACCTGTCGCCTACCTGTCAAAGAAGCT 3709

 517 uAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaAlaV 534
     ||||||||||||||||||||||||||:::|||||||||||||||||||||
3710 CGATCCTGTAGCCAGTGGTTGGCCCATATGCCTGAAGGCTATCGCAGCTG 3759

 534 alAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIle 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3760 TGGCCATACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAGAATATA 3809

 551 ThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAs 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3810 ACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAGA 3859

 567 pArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuT 584
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3860 CCGATGGATGACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTCA 3909

 584 hrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3910 CAGAGAGGGTCACGTTCGCTCCACCAGCCGCTCTCAACCCTGCCACTCTT 3959

 601 LeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeuLe 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3960 CTGCCTGAAGAGACTGATGAACCAGTGACTCATGATTGCCATCAACTATT 4009

 617 uIleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThrG 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||
4010 GATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTG 4059

 634 lyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGluGly 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
4060 GAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGGT 4109

 651 LysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAl 667
     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
4110 AAGAGGATGGCTGGGGCGGCGGTGGTGGACGGGACCCGCACGATCTGGGC 4159

 667 aSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaL 684
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4160 CAGCAGCCTGCCGGAAGGAACTTCAGCACAAAAGGCTGAGCTCATGGCCC 4209

 684 euThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4210 TCACGCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAAACATTTATACG 4259

 701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrLy 717
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4260 GACAGCAGGTATGCCTTTGCGACTGCACACGTACATGGGGCCATCTATAA 4309

 717 sGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluG 734
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4310 ACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAGG 4359

 734 luIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIle 750
     |||||||||||||||||||||||:::|||||||||||||||||||||||||
4360 AAATTCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCTATT 4409

 751 IleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAs 767
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4410 ATACACTGTCCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAA 4459

 767 nGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeuL 784
     |||||||||||||||||||||||||||||||||||:::||||||||||||
4460 CCAGATGGCTGACCGGGTTGCCAAGCAGGCAGCCCAGGGTGTTAACCTTC 4509

 784 euProIleIleGluThrProLysAlaProGluProArgArgGlnTyrThr 800
     |||||||||||||   |||||||||||||||||||||||||||||||||||
4510 TGCCTATAATAGAAATGCCCAAAGCCCCAGAACCCAGACGACAGTACACC 4559

 801 LeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrPr 817
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4560 CTAGAAGACTGGCAAGAGATAAAAAAGATAGACCAGTTCTCTGAGACTCC 4609

 817 oGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysG 834
     ||||||||||||||||||||||   |||||||||||||||||||||||||||
4610 GGAAGGGACCTGCTATACCTCAGATGGGAAGGAAATCCTGCCCCACAAAG 4659

 834 luGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThr 850
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4660 AAGGGTTAGAATATGTCCAACAGATACATCGTCTAACCCACCTAGGAACT 4709

 851 LysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLe 867
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4710 AAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGGCT 4759

 867 uProGlyValAlaAspSerValValLysHisCysValProCysGlnLeuV 884
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4760 ACCAGGAGTGGCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGG 4809

 884 alAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySer 900
     |||||||||||||||||||:::|||||||||||||||||||||||||||||
4810 TTAATGCTAATCCTTCCAGAATGCCTCCAGGGAAGAGACTAAGGGGAAGC 4859

 901 HisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLy 917
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4860 CACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAA 4909

 917 sTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpV 934
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4910 ATACGGAAACAAATACCTATTGGTTTTTGTAGACACCTTTTCAGGATGGG 4959

 934 alGluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLys 950
     |||||||||||||||||||||||||||||||||||||||||||||||||||
4960 TAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAAAAA 5009

 951 IleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySe 967
     |||||||||||||||||||||||||||||||||||||||||||||||||||
5010 ATACTGGAAGAAATTTTTCCAAGATTTGGAATACCTAAGGTAATAGGGTC 5059

 967 rAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysI 984
     |||||||||||||||||||||||||||||||||||||||||||||||||||
5060 AGACAATGGTCCAGCTTTTGTTGCCCAGGTAAGTCAGGGACTGGCCAAGA 5109

 984 leLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSer 1000
     |||||||||||||||||||||||||||||||||||||||||||||||||||
5110 TATTGGGGATTGATTGGAAACTGCATTGTGCATACAGACCCCAAAGCTCA 5159

1001 GlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLe 1017
     |||||||||||||||||||||||||||||||||||||||||||||||||||
5160 GGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACCCTTACTAAATT 5209

1017 uThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValL 1034
     ||||:::|||||||||:::|||||||||:::|||||||||||||||||||
5210 GACCGCGGAGACTGGCGTTAATGATTGGATAGCTCTCCTGCCCTTTGTGC 5259

1034 euPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGlu 1050
     |||||||||||||||||||||||||||||||||||||||||||||||||||
5260 TTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAA 5309

1051 LeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisSe 1067
     |||||||||||||||||||||||||||:::|||||||||   :::|||||
5310 TTACTCTACGGGGGACCCCCCCCCATTGGTAGAAATTGCTTCTGTACATAG 5359

1067 rAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuG 1084
     |||||||||||||||||||||||||||||||||||||||||||||||||||
5360 TGCTGACGTGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGCACTTG 5409

1084 luTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSer... 1099
     |||||||||||||||||||||||||:::|||||||||||||||||||||
5410 AGTGGGTGAGACAACGAGCGTGGAGGCAACTCCGGGAGGCCTACTCAGGA 5459

1100 GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerValTy 1116
     ||||||||||||||||:::|||||||||||||||||||||||||||||||
5460 GGAGGAGACTTGCAGATCCCACATCGTTTCCAAGTGGGAGATTCAGTCTA 5509

1116 rValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProT 1133
     |||||||||||||||||||||||||||||||||||||||||||||||||||
5510 CGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGCCCTT 5559

1133 yrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIlePro 1149
     |||||||||||||||||||||||||||||||||||||||||||||||||||
5560 ATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCC 5609

1150 ThrTrpIleHisAlaSerHisValLysProAlaProProProAspSerGl 1166
     |||||||||||||||||||||||||||||||||||||||||||||||||||
5610 ACCTGGATCCATGCATCCCACGTTAAACCGGCGCCACCTCCCGATTCGGG 5659

1166 yTrpArgAlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgL 1183
     ||||:::|||||||||||||||||||||||||||||||||||||||||||:
5660 GTGGAAAGCCGAAAAGACTGAAAATCCCCTTAAGCTTCGCCTCCATCGCG 5709

1183 euValProTyrSerAsnAsnAsn 1190
     ::|||||||||||   ||||||
5710 TGGTTCCTTACTCTGTCAATAAC 5732
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-1

```
seq_documentation_block:
; Sequence 1, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
```

```
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1


alignment_scores:
            Quality: 5637.50              Length:   1127
              Ratio:  5.061                Gaps:      3
 Percent Similarity:  98.846   Percent Identity: 96.628

alignment_block:
 US-09-171-553B-5 x US-08-766-528-1  ...

 Align seg 1/1  to: US-08-766-528-1  from: 1  to: 8060

       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4687 GGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGT 4736

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisServ 34
         ||||||||||||||||||||||||||||||||||||||:::|||||||
    4737 GGAGGGGCAACCAGTTGAGTTCCTGGTTGATACCGGAGCGAAACATTCAG 4786

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
         ||||||||||||||||||||||||||||:::|||||||||||||||||||
    4787 TGCTACTACAGCCATTAGGAAAACTAAAAGATAAAAAATCCTGGGTGATG 4836

      51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
         |||   |||||||||:::|||||||||||||||||||||   ||||||||
    4837 GGTGC.ACAGGGCAACAACAGTATCCATGGACTACCCGAAG.ACAGTTGA 4884

      67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4885 CTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTCATACCTGAGTGCC 4934

      84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
         ||:::|||||||||||||||||||||||||||||||||||||||||||||
    4935 CAGCACCCCTCTTAGGTAGAGACTTATTGACCAAGATGGGAGCACAAATT 4984

     101 SerPheGluGlnGlyArgProGluValSerValAsnAsnLysProIleTh 117
         ||||||||||||||:::|||||||||||||:::|||||||||||||||||
    4985 TCTTTTGAACAAGGGAAACCAGAAGTGTCTGCAAATAACAAACCTATCAC 5034

     117 rValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnV 134
         ||||||||||||||||||||||||||||||||||||||||||||||   |
    5035 TGTGTTGACCCTCCAATTAGATGACGAATATCGACTATACTCTCCCCTAG 5084

     134 alLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGlnAla 150
         ||||||||||||||:::|||||||   |||||||||||||||||||||||

    5085 TAAAGCCTGATCAAAATATACAATTCTGGTTGGAACAGTTTCCCCAAGCC 5134

     151 TrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5135 TGGGCAGAAACCGCAGGGATGGGTTTGGCAAAGCAAGTTCCCCCACAAGT 5184

     167 lIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrProL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5185 TATTCAACTGAAGGCCAGTGCCACACCAGTGTCAGTCAGACAGTACCCCT 5234

     184 euSerArgGluAlaArgGluGlyIleTrpProHisValGlnArgLeuIle 200
         |||||:::||||||:::|||||||||   ||||||||||||||||||||
    5235 TGAGTAAAGAAGCTCAAGAAGGAATTCGGCCGCATGTCCAAAGATTAATC 5284

     201 GlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLe 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5285 CAACAGGGCATCCTAGTTCCTGTCCAATCTCCCTGGAATACTCCCCTGCT 5334

     217 uProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeuA 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5335 ACCGGTTAGAAAGCCTGGGACTAATGACTATCGACCAGTACAGGACTTGA 5384

     234 rgGluValAsnLysArgValGlnAspIleHisProThrValProAsnPro 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5385 GAGAGGTCAATAAACGGGTGCAGGATATACACCCAACAGTCCCGAACCCT 5434

     251 TyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLe 267
         ||||||||||||   ||||||||||||:::|||:::|||||||||||||||
    5435 TATAACCTCTTGTGTGCTCTCCCACCCCAACGGAGCTGGTATACAGTATT 5484

     267 uAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnP 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5485 GGACTTAAAGGATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAAC 5534

     284 roLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGln 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5535 CACTTTTTGCCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGGCAG 5584

     301 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePh 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5585 CTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTT 5634

     317 eAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5635 TGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATCCAACACCCTC 5684

     334 lnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAlaThr 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5685 AGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACC 5734

     351 LysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAs 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5735 AAACAGGACTGCTTAGAAGGCACGAAGGCACTACTGCTGGAATTGTCTGA 5784

     367 pLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluV 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5785 CCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGG 5834

     384 alThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrGlu 400
         ||||||||||||||||||||||||||   ||||||||||||||||||||||
    5835 TAACATACTTGGGGTACAGTTTACGGGACGGGCAGCGATGGCTGACGGAG 5884

     401 AlaArgLysLysThrValValGlnIleProAlaProThrThrAlaLysGl 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5885 GCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACA 5934

     417 nValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProG 434
         |:::||||||||||||||||||||||||||||||||||||||||||||||||
    5935 AATGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGG 5984

     434 lyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGly 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5985 GGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGG 6034
```

451 GlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLy 467
6035 GAATTCTCCTGGGCTCCTGAGCACCAGAAGGCATTTGATGCTATCAAAAA 6084
467 sAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLysProP 484
6085 GGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCT 6134
484 heThrLeuTyrValAspGluArgLysGlyValAlaArgGlyValLeuThr 500
6135 TTACCCTTTATGTGGATGAGCGTAAGGGAGTAGCCCGGGGAGTTTTAACC 6184
501 GlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLe 517
6185 CAAACCCTAGGACCATGGAGAAGACCTGTCGCCTACCTGTCAAAGAAGCT 6234
517 uAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaAlaV 534
6235 CGATCCTGTAGCCAGTGGTTGGCCCATATGCCTGAAGGCTATCGCAGCTG 6284
534 alAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln.AsnIl 550
6285 TGGCCATACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAAGAATAT 6334
550 eThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProA 567
6335 AACTGTAATAGCCCCCCATGCATTGGAGAACATCGTTCGGCAGCCCCCAG 6384
567 spArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeu 583
6385 ACCGATGGATGACCAACGCCCGCATGACCCACTATCAAAGCCTGCTTCTC 6434
584 ThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLe 600
6435 ACAGAGAGGGTCACGTTCGCTCCACCAACCGCTCTCAACCCTGCCACTCT 6484
600 uLeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeuL 617
6485 TCTGCCTGAAGAGACTGATGAACCAGTGACTCATGATTGCCATCAACTAT 6534
617 euIleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThr 633
6535 TGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACT 6584
634 GlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGluGl 650
6585 GGAGAAGTGCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAGG 6634
650 yLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpA 667
6635 TAAGAGGATGGCTGGGGCGGCGGTGGTGGACGGGACCCGCACGATCTGGG 6684
667 laSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAla 683
6685 CCAGCAGCCTGCCGGGAGGAACTTCAGCACAAAAGGCTGAGCTCATGGCC 6734
684 LeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrTh 700
6735 CTCACGCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAAACATTTATAC 6784
700 rAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrL 717
6785 GGACAGCAGGTATGCCTTTGCGACTGCACACGTACATGGGGCCATCTATA 6834
717 ysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGlu 733
6835 AACAAAGGGGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAG 6884
734 GluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIl 750
6885 GAAATTCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCTAT 6934
750 eIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyA 767
6935 TATACACTGTCCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAA 6984
767 snGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeu 783
6985 ACCAGATGGCTGACCGGGTTGCCAAGCAGGCAGCCCAGGGTGTTAACCTT 7034
784 LeuProIleIleGluThrProLysAlaProGluProArgArgGlnTyrTh 800
7035 CTGCCTATAATAGAAATGCCCAAAGCCCCAGAACCCAGACGACAGTACAC 7084
800 rLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrP 817
7085 CCTAGAAGACTGGCAAGAGATAAAAAAGATAGACCA.TTCTCTGAGACTC 7133
817 roGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLys 833
7134 CGGAAGGGACCTGCTATACCTCAGATGGGAAGGAAATCCTGCCCCACAAA 7183
834 GluGlyLeuGluTyrValGlnGln.IleHisArgLeuThrHisLeuGlyT 850
7184 GAAGGGTTAGAATATGTCCAACAAGATACATCGTCTAACCCACCTAGGAA 7233
850 hrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArg 866
7234 CTAAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGG 7283
867 LeuProGlyValAlaAspSerValValLysHisCysValProCysGlnLe 883
7284 CTACCAGGAGTGGCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCT 7333
883 uValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlyS 900
7334 GGTTAATGCTAATCCTTCCAGAATGCCTCCAGGGAAGAGACTAAGGGGAA 7383
900 erHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAla 916
7384 GCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCT 7433
917 LysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTr 933
7434 AAATATGGAAACAAATACCTATTGGTTTTTGTAGACACCTTTTCAGGATG 7483
933 pValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLysL 950
7484 GGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTAGCTAAAA 7533
950 ysIleLeuGluGluIleProArgPheGlyIleProLysValIleGly 966
7534 AAATACTGGAAGAAATTTTTCCAAGATTTGGAATACCTAAGGTAATAGGG 7583
967 SerAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLy 983
7584 TCAGACAATGGTCCAGCTTTTGTTGCCCAGGTAAGTCAGGGACTGGCCAA 7633
983 sIleLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerS 1000
7634 GATATTGGGGATTGATTGGAAACTGCATTGTGCATACAGACCCCAAAGCT 7683
1000 erGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLys 1016
7684 CAGGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACCCTTACTAAA 7733
1017 LeuThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVa 1033
7734 TTGACCGCGGAGACTGGCGTTAATGATTGGATAGCTCTCCTGCCCTTTGT 7783
1033 lLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrG 1050
7784 GCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATG 7833
1050 luLeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHis 1066

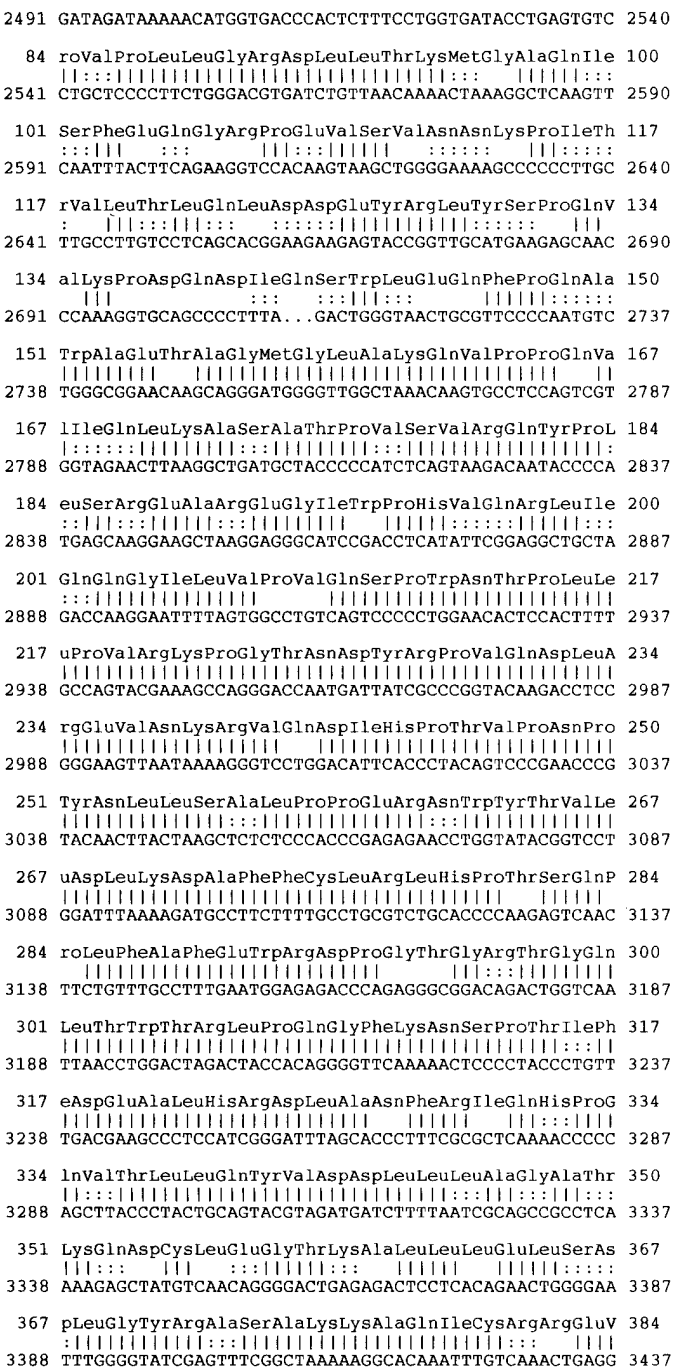
```
                                                                  :::|||
     7834 AATTACTCTACGGGGGACCCCCCCCCATTGGTAGAAATTGCTTCTGTACAT 7883

     1067 SerAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLe 1083
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     7884 AGTGCTGATGTGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGCACT 7933

     1083 uGluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSer. 1099
          ||||||||||||||||||||||||:::||||||||||||||||||||||
     7934 TGAGTGGGTGAGACAACGAGCGTGGAGGCAACTCCGGGAGGCCTACTCAG 7983

     1100 ..GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspServVal 1115
            ||||||||||||||:::||||||||||||||||||||||||||||||||
     7984 GAGGAGGAGACTTGCAGATCCCACATCGTTTCCAAGTGGGAGATTCAGTC 8033

     1116 TyrValArgArgHisArgAlaGlyAsn 1124
          ||||||||||||||||||||||||||
     8034 TACGTTAGACGCCACCGTGCAGGAAAC 8060

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-2

seq_documentation_block:
; Sequence 2, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;     APPLICANT:  Jay A. Fishman
;     TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;     TITLE OF INVENTION:  AND METHODS OF USE
;     NUMBER OF SEQUENCES:  74
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;       STREET:  60 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109-1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/766,528
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/572,645
;       FILING DATE:  14-DEC-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Louis Myers
;       REGISTRATION NUMBER:  35,965
;       REFERENCE/DOCKET NUMBER:  MGP-038CP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227-5941
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7333 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-766-528-2


alignment_scores:
            Quality: 4489.50              Length:   1191
              Ratio:  5.033                Gaps:      2
 Percent Similarity:  74.895   Percent Identity: 74.643

alignment_block:
 US-09-171-553B-5 x US-08-766-528-2   ..

 Align seg 1/1  to: US-08-766-528-2  from: 1  to: 7333
```

```
       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2173 GGGAGACGGGGTTCGGACCCCCTCCCCGAGCCCAGGGTAACTTTGAAGGT 2222

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerV 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2223 GGAGGGGCAACCAGTTGAGTTCCTGGTTGATACCGGAGCGGAGCATTCAG 2272

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2273 TGCTGCTACAACCATTAGGAAAACTAAAAGAAAAAAAATCCTGGGTGATG 2322

      51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2323 GGTGCCACAGGGCAACGGCAGTATCCATGGACTACCCGAAGAACCGTTGA 2372

      67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2373 CTTGGGAGTGGGACGGGTAACCCACTCGTTTCTGGTCATCCCTGAGTGCC 2422

      84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2423 CAGTACCCCTTCTAGGTAGAGACTTACTGACCAAGATGGGAGCTCAAATT 2472

     101 SerPheGluGlnGlyArgProGluValSerValAsnAsnLysProIleTh 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2473 TCTTTTGAACAAGGAAGACCAGAAGTGTCTGTGAATAACAAACCCATCAC 2522

     117 rValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnV 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2523 TGTGTTGACCCTCCAATTAGATGATGAATATCGACTATATTCTCCCCAAG 2572

     134 alLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGlnAla 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2573 TAAAGCCTGATCAAGATATACAGTCCTGGTTGGAGCAGTTTCCCCAAGCC 2622

     151 TrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnVa 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2623 TGGGCAGAAACCGCAGGGATGGGTTTGGCAAAGCAAGTTCCCCCACAGGT 2672

     167 lIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrProL 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2673 TATTCAACTGAAGGCCAGTGCTACACCAGTATCAGTCAGACAGTACCCCT 2722

     184 euSerArgGluAlaArgGluGlyIleTrpProHisValGlnArgLeuIle 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2723 TGAGTAGAGAGGCTCGAGAAGGAATTTGGCCGCATGTTCAAAGATTAATC 2772

     201 GlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLe 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2773 CAACAGGGCATCCTAGTTCCTGTCCAATCCCCTTGGAATACTCCCCTGCT 2822

     217 uProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeuA 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2823 ACCGGTTAGGAAGCCTGGGACCAATGATTATCGACCAGTACAGGACTTGA 2872

     234 rgGluValAsnLysArgValGlnAspIleHisProThrValProAsnPro 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2873 GAGAGGTCAATAAAAGGGTGCAGGACATACACCCAACGGTCCCGAACCCT 2922

     251 TyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2923 TATAACCTCTTGAGCGCCCTCCCGCCTGAACGGAACTGGTACACAGTATT 2972

     267 uAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnP 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2973 GGACTTAAAAGATGCCTTCTTCTGCCTGAGATTACACCCCACTAGCCAAC 3022

     284 roLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGln 300
         ||||||||:::||||||||||||||||||||||||||||||||||||||||
    3023 CACTTTTTACCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGGCAG 3072
```

```
 301 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePh 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3073 CTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTT 3122

 317 eAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3123 TGACGAAGCCCTACACAGGGACCTGGCCAACTTCAGGATCCAACACCCTC 3172

 334 lnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAlaThr 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3173 AGGTGACCCTCCTCCAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACC 3222

 351 LysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAs 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3223 AAACAGGACTGCTTAGAAGGTACGAAGGCACTACTGCTGGAATTGTCTGA 3272

 367 pLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluV 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3273 CCTAGGCTACAGAGCCTCTGCTAAGAAGGCCCAGATTTGCAGGAGAGAGG 3322

 384 alThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrGlu 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3323 TAACATACTTGGGGTACAGTTTGCGGGGCGGGCAGCGATGGCTGACGGAG 3372

 401 AlaArgLysLysThrValValGlnIleProAlaProThrThrAlaLysGl 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3373 GCACGGAAGAAAACTGTAGTCCAGATACCGGCCCCAACCACAGCCAAACA 3422

 417 nValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProG 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3423 AGTGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGG 3472

 434 lyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGly 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3473 GGTTTGCGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAGAAAAAGGG 3522

 451 GlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLy 467
     |||
3523 GGTTG..................................................  3527

 467 sAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLysProP 484

3527 ..................................................  3527

 484 heThrLeuTyrValAspGluArgLysGlyValAlaArgGlyValLeuThr 500

3527 ..................................................  3527

 501 GlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLe 517

3527 ..................................................  3527

 517 uAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaAlaV 534

3527 ..................................................  3527

 534 alAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIle 550

3527 ..................................................  3527

 551 ThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAs 567

3527 ..................................................  3527

 567 pArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuT 584

3527 ..................................................  3527

 584 hrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600

3527 ..................................................  3527

 601 LeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeuLe 617

3527 ..................................................  3527

 617 uIleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThrG 634

3527 ..................................................  3527

 634 lyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGluGly 650

3527 ..................................................  3527

 651 LysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAl 667

3527 ..................................................  3527

 667 aSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaL 684

3527 ..................................................  3527

 684 euThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700

3527 ..................................................  3527

 701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrLy 717

3527 ..................................................  3527

 717 sGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluG 734
                  ||||||||||||||||||||||||||||||||||||
3528 .............CTTACCTCAGCAGGGAGGGAAATAAAGAACAAAGAGG 3564

 734 luIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIle 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3565 AAATTCTAAGCCTATTAGAAGCCTTACATTTGCCAAAAAGGCTAGCTATT 3614

 751 IleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAs 767
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3615 ATACACTGTCCTGGACATCAGAAAGCCAAAGATCTCATATCTAGAGGGAA 3664

 767 nGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeuL 784
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3665 CCAGATGGCTGACCGGGTTGCCAAGCAGGCAGCCCAGGCTGTTAACCTTC 3714

 784 euProIleIleGluThrProLysAlaProGluProArgArgGlnTyrThr 800
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3715 TGCCTATAATAGAAACGCCCAAAGCCCCAGAACCCAGACGACAGTACACC 3764

 801 LeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrPr 817
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3765 CTAGAAGACTGGCAAGAGATAAAAAAGATAGACCAGTTCTCTGAGACTCC 3814

 817 oGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysG 834
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3815 GGAGGGGACCTGCTATACCTCATATGGGAAGGAAATCCTGCCCCACAAAG 3864

 834 luGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThr 850
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3865 AAGGGTTAGAATATGTCCAACAGATACATCGTCTAACCCACCTAGGAACT 3914

 851 LysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLe 867
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3915 AAACACCTGCAGCAGTTGGTCAGAACATCCCCTTATCATGTTCTGAGGCT 3964

 867 uProGlyValAlaAspSerValValLysHisCysValProCysGlnLeuV 884
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
3965 ACCAGGAGTGGCTGACTCGGTGGTCAAACATTGTGTGCCCTGCCAGCTGG 4014

 884 alAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySer 900
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
4015 TTAATGCTAATCCTTCCAGAATACCTCCAGGAAAGAGACTAAGGGGAAGC 4064

 901 HisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLy 917
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 4065 CACCCAGGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAA 4114

  917 sTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpV 934
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 4115 ATACGGAAACAAATATCTATTGGTTTTTGTAGACACCTTTTCAGGATGGG 4164

  934 alGluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLys 950
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 4165 TAGAGGCTTATCCTACTAAAAAAGAGACTTCAACCGTGGTGGCTAAGAAA 4214

  951 IleLeuGluGluIleProArgPheGlyIleProLysValIleGlySe 967
      ||||||||||||||||||||||||||||||||||||||||||||||||
 4215 ATACTGGAGGAAATTTTTCCAAGATTTGGAATACCTAAGGTAATAGGGTC 4264

  967 rAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysI 984
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 4265 AGACAATGGTCCAGCTTTCGTTGCCCAGGTAAGTCAGGGACTGGCCAAGA 4314

  984 leLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSer 1000
      ||||||||||||||   ||||||||||||||||||||||||||||||||
 4315 TATTGGGGATTGATTGAAAACTGCATTGTGCATACAGACCCCAAAGCTCA 4364

 1001 GlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLe 1017
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 4365 GGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACCCTTACCAAATT 4414

 1017 uThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValL 1034
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 4415 GACCACAGAGACTGGCATTAATGATTGGATGGCTCTCCTGCCCTTTGTGC 4464

 1034 euPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGlu 1050
      ||||||||||||||||||||||||||||||||||||||||||||||:::
 4465 TTTTTAGGGTGAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATAAA 4514

 1051 LeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisSe 1067
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 4515 TTGCTCTACGGGGGACCCCCCCCGTTGGCAGAAATTGCCTTTGCACATAG 4564

 1067 rAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuG 1084
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 4565 TGCTGATGTGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGCGCTCG 4614

 1084 luTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGly 1100
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 4615 AGTGGGTGAGGCAGCGAGCGTGGAAGCAGCTCCGGGAGGCCTACTCAGGA 4664

 1101 GlyAspLeuGlnValProHisArgPheGlnValGlyAspSerValTyrVa 1117
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 4665 GGAGACTTGCAAGTTCCACATCGCTTCCAAGTTGGAGATTCAGTCTATGT 4714

 1117 lArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProTyrL 1134
      ||||||||||||||||||||||||||||||||||   ||||||||||||||
 4715 TAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTAGAAGGGACCTTATC 4764

 1134 euValLeuLeuThrThrProThrAlaValLysValGluGlyIleProThr 1150
      ||||||||||||||||||||||||||||||||||||
 4765 TCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGG.......... 4804

 1151 TrpIleHisAlaSerHisValLysProAlaProProProAspSerGlyTr 1167

 4804 .................................................. 4804

 1167 pArgAlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgLeuV 1184
                        ||||||||||||||||||||||||||||:::|
 4805 ..................AATCCCCTTAAGCTTCGCCTCCATCGCGTGG 4835

 1184 alProTyrSerAsnAsnAsnSer 1191
      ||||||||||   |||||||||
 4836 TTCCTTACTCTGTCAATAACTCC 4858

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-075-272-1

seq_documentation_block:
```

```
; Sequence 1, Application US/09075272
; Patent No. 6136598
;  GENERAL INFORMATION:
;    APPLICANT:  MILLER, A. DUSTY
;    APPLICANT:  WOLGAMOT, GREG
;    APPLICANT:  BONHAM, LYNN
;    TITLE OF INVENTION:  MUS DUNNI ENDOGENOUS RETROVIRAL
;    TITLE OF INVENTION:  PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,272
;      FILING DATE:  08-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/046,140
;      FILING DATE:  09-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poor, Brian W.
;      REGISTRATION NUMBER:  32,928
;      REFERENCE/DOCKET NUMBER:  14538A-003710
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-075-272-1


alignment_scores:
            Quality: 4339.50              Length:   1202
              Ratio:  4.201                Gaps:      8
 Percent Similarity:  85.940   Percent Identity: 68.968

alignment_block:
 US-09-171-553B-5 x US-09-075-272-1   ..

 Align seg 1/1  to: US-09-075-272-1  from: 1  to: 8655

       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         |||   ||||||||||||||||||||||||||||||||||||||:::||
    2291 GGAAGTCGGGGCTCAGACCCCCTCCCTGAGCCTAGGGTAACTTTGTCTGT 2340

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisServ 34
         |||||||   ||||||:::||||||:::||||||||||||||||||||
    2341 GGAGGGGACTCCCGTTAATTTCCTGATAGATACTGGAGCAGAGCATTCGG 2390

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
         |||||   :::|||||||||||||||   :::|||:::::   |||:::
    2391 TACTCACCAGCCCCCTGGGCAAACTAGGCTCTAAAAGGACCATAGTGGTT 2440

      51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
         ||||||||||||::::::   |||||||||||||||:::|||:::::::
    2441 GGAGCCACTGGGAGTAAACTTTACCCCTGGACAACCAAAAGAGCCTTACA 2490

      67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
         ::::      :::   ||||||||||||||||||||||||||||||||||
```


2491 GATAGATAAAAACATGGTGACCCACTCTTTCCTGGTGATACCTGAGTGTC 2540
84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
2541 CTGCTCCCCTTCTGGGACGTGATCTGTTAACAAAACTAAAGGCTCAAGTT 2590
101 SerPheGluGlnGlyArgProGluValSerValAsnAsnLysProIleTh 117
2591 CAATTTACTTCAGAAGGTCCACAAGTAAGCTGGGGAAAAGCCCCCCTTGC 2640
117 rValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnV 134
2641 TTGCCTTGTCCTCAGCACGGAAGAAGAGTACCGGTTGCATGAAGAGCAAC 2690
134 alLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGlnAla 150
2691 CCAAAGGTGCAGCCCCTTTA...GACTGGGTAACTGCGTTCCCCAATGTC 2737
151 TrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnVa 167
2738 TGGGCGGAACAAGCAGGGATGGGGTTGGCTAAACAAGTGCCTCCAGTCGT 2787
167 lIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrProL 184
2788 GGTAGAACTTAAGGCTGATGCTACCCCCATCTCAGTAAGACAATACCCCA 2837
184 euSerArgGluAlaArgGluGlyIleTrpProHisValGlnArgLeuIle 200
2838 TGAGCAAGGAAGCTAAGGAGGGCATCCGACCTCATATTCGGAGGCTGCTA 2887
201 GlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLe 217
2888 GACCAAGGAATTTTAGTGGCCTGTCAGTCCCCCTGGAACACTCCACTTTT 2937
217 uProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeuA 234
2938 GCCAGTACGAAAGCCAGGGACCAATGATTATCGCCCGGTACAAGACCTCC 2987
234 rgGluValAsnLysArgValGlnAspIleHisProThrValProAsnPro 250
2988 GGGAAGTTAATAAAAGGGTCCTGGACATTCACCCTACAGTCCCGAACCCG 3037
251 TyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLe 267
3038 TACAACTTACTAAGCTCTCTCCCACCCGAGAGAACCTGGTATACGGTCCT 3087
267 uAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnP 284
3088 GGATTTAAAAGATGCCTTCTTTTGCCTGCGTCTGCACCCCAAGAGTCAAC 3137
284 roLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGln 300
3138 TTCTGTTTGCCTTTGAATGGAGAGACCCAGAGGGCGGACAGACTGGTCAA 3187
301 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePh 317
3188 TTAACCTGGACTAGACTACCACAGGGGTTCAAAAACTCCCCTACCCTGTT 3237
317 eAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
3238 TGACGAAGCCCTCCATCGGGATTTAGCACCCTTTCGCGCTCAAAACCCCC 3287
334 lnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAlaThr 350
3288 AGCTTACCCTACTGCAGTACGTAGATGATCTTTTAATCGCAGCCGCCTCA 3337
351 LysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAs 367
3338 AAAGAGCTATGTCAACAGGGGACTGAGAGACTCCTCACAGAACTGGGGAA 3387
367 pLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluV 384
3388 TTTGGGGTATCGAGTTTCGGCTAAAAAGGCACAAATTTGTCAAACTGAGG 3437
384 alThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrGlu 400
3438 TAATCTACTTGGGGTATACTTTGCGAGGAGGTAAAAGATGGCTCACTGAA 3487
401 AlaArgLysLysThrValValGlnIleProAlaProThrThrAlaLysGl 417
3488 GCCCGGAAAAAGACTGTCATGATGATTCCACCGCCAACCACACCACGGCA 3537
417 nValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProG 434
3538 GGTACGTGAGTTTCTGGGGACTGCTGGCTTTTGTAGACTCTGGATTCCAG 3587
434 lyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGly 450
3588 GCTTTGCGACCCTAGCAGCACCCCTATATCCCTTGACTAGGGAAGGAATT 3637
451 GlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLy 467
3638 CCCTTTGAATGGAAAGAAGAACACCAAAGAGCTTTTGAGGCTATCAAATC 3687
467 sAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLysProP 484
3688 GTCTCTAATGACTGCCCCTGCGCTAGCATTACCAGACTTGACTAAGTCCT 3737
484 heThrLeuTyrValAspGluArgLysGlyValAlaArgGlyValLeuThr 500
3738 TCGTCCTATATGTGGACGAGAGAGCGGGCATAGCCAGAGGGGTGCTGACA 3787
501 GlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLe 517
3788 CAAGCACTAGGACCCTGGAAGAGACCGGTAGCCTATCTGTCAAAAAAACT 3837
517 uAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaAlaV 534
3838 GGATCCCGTTGCTAGTGGATGGCCCACATGTCTGAAAGCTATTGCAGCAG 3887
534 alAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIle 550
3888 TAGCCCTGCTGATCAAAGATGCTGATAAACTGACAATGGGACAGCAGGTG 3937
551 ThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAs 567
3938 ACTGTCGTGGCCCCTCATGCCCTGGAGAGTATTGTGCGGCAGCCACCTGA 3987
567 pArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuT 584
3988 TAGATGGATGACTAATGCCCGGATGACACACTATCAGAGTCTGCTGCTAA 4037
584 hrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600
4038 ATGACCGGGTAACCTTTGCCCCCCCTGCCATTCTCAACCCAGCCACCCTC 4087
601 LeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeuLe 617
4088 CTCCCTCTAACGAATGATTCCGTCCCAGTACATCGATGTGCAGACATCCT 4137
617 uIleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThrG 634
4138 GGCTGAAGAAATTGGGACCAGAAAAGACCTGACTGACCAACCCTGGCCTG 4187
634 lyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGluGly 650
4188 GA...GCTCCTAGCTGGTACACGGATGGCAGCAGTTTTCTGATTGAGGGA 4234
651 LysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAl 667
4235 AAGCGAAGGGCTGGAGCTGCGGTGGTGGATGGAAAAAAGGTAATTTGGGC 4284
667 aSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaL 684
4285 AAGTGCCTTGCCTGAAGGAACTTCGGCACAGAAGGCTGAACTCATAGCGC 4334

```
 684 euThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700
     ||||||||||||||||   ||||||||||||   ||||||||||||||||
4335 TTACACAGGCCCTCCGAGAGGCCGAAGGTAAGATCATTAATATTTACACT 4384

 701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrLy 717
     ||||||||||||||||||||||||||||:::|||||||||||||||::
4385 GACAGCCGCTATGCTTTTGCTACTGCACATATCCATGGGGCCATCTACAG 4434

 717 sGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluG 734
     :|||||||||||||||||||||||||||:::::::|||||||||||||||
4435 GCAGCGAGGGTTATTGACTTCAGCCGGTAAAGACATCAAAAACAAAGAAG 4484

 734 luIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIle 750
     ||||||||:::||||||||||||:::|||   ||||||::::::||||||
4485 AAATTCTGGCCCTGTTAGAAGCCATACATGCGCCCAAGAAGGTAGCCATC 4534

 751 IleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAs 767
     ||||||||||||||||||||||||:::::::||||||::::::::||||||
4535 ATCCACTGTCCCGGTCACCAGAAAGGAGAGGACTTAGTGGCCAAGGGCAA 4584

 767 nGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeuL 784
     |:::|||||||||   ||||||||||||:::||||||::::::   :::|
4585 CCGGATGGCAGACTCAGTAGCAAAACAGGTTGCTCAGGGGGCCATGATCT 4634

 784 euProIleIleGluThrPro...LysAlaProGluProArgArgGlnTyr 799
     ||            :::|||   |||:::||||||
4635 TAACTGAAAAAGGAAATCCGTCCAAAAGCCCTGAG..............  4669

 800 ThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluTh 816
                 :::|||       |||   ::::::      ||
4670 ..............GATGAAAACTATGATATAAAAGAACTATTTTGGAC 4704

 816 rProGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisL 833
     |   :::         |||        ||||||      |||      :
4705 TAGTGATCCCCTCCCATACTTTTTCGAAGGAAAAATAGACTTGACACCCG 4754

 833 ysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 849
     ::|||||:::::::::|||:::   :::|||::::::||||||||||||
4755 AGGAAGGAATAAAATTTGTGAAAGGACTACACCAGTTTACCCACCTGGGA 4804

 850 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuAr 866
     ::::::   ::::::::::|||::::::   |||   |||:::|||   ::
4805 GTTGAAAAAATGATGAGACTGATTAAAAAGTCCCGGTATCAAGTCCCTAA 4854

 866 gLeuProGlyValAlaAspSerValValLysHisCysValProCysGlnL 883
     :|||   :::|||||||:::::::::::::::   |||      |||   :
4855 CTTGAAGTCAGTGGCTCAAAAGATTATAAACTCCTGCAAAGCGTGTGCAT 4904

 883 euValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGly 899
     :::::|||||:::   :::       ||||||||||||   ||||||
4905 TCACTAATGCAACTAAAACCTACAAAGAACCTGGAAAGAGACAACGGGGA 4954

 900 SerHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAl 916
     ::::::||||||::::::||||||||||||||||||||||||||||||::
4955 GACCGTCCTGGAGTGTATTGGGAGGTGGATTTTACTGAAGTTAAACCCGG 5004

 916 aLysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyT 933
     :   ||||||||||||||||||||||||||||||||||||||||||||||
5005 AATGTATGGTAATAAGTATCTGTTAGTATTTGTAGACACCTTTTCAGGAT 5054

 933 rpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 949
     ||||||||||:::|||||||||   ||||||:::   :::|||||||||
5055 GGGTTGAAGCGTTCCCCACTAAAACTGAGACTGCCCAGATTGTGGCCAAG 5104

 950 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGl 966
     ||||||:::|||||||||:::|||||:::|||:::|||||||||||||||
5105 AAGATTTTTGAAGAAATCCTGCCAAGATATGGAGTACCTAAGGTAATCGG 5154

 966 ySerAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaL 983
     ||||||||||||||||||||||||||||||||||||||||||||||||||
5155 GTCCGACAATGGACCAGCCTTTGTTGCCCAGGTAAGTCAGGGATTGGCCA 5204

 983 ysIleLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSer 999
```

```
                ||||||||||||||||||||||||||||||||||||||||||||
5205 CTCAGTTGGGCATTGATTGGAAATTACATTGTGCTTACCGCCCTCAAAGC 5254

1000 SerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLy 1016
     ||||||||||||||||||||||||||||:::|||||||||||||||||||
5255 TCAGGACAGGTAGAGAGGATGAATAGAACATTAAAGGAGACCTTGACTAA 5304

1016 sLeuThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheV 1033
     ||||:::   |||||||||   :::||||||:::||||||||||||:::::
5305 ATTGGCCATGGAGACCGGCGGGAAAGACTGGGTGGCTCTCCTCCCCCTTG 5354

1033 alLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1049
     ::||||||||:::||||||||||||||||:::|||||||||||||||:::
5355 CGCTCTTCCGAGCCCGGAACACCCCCGGACGTTTTGGGCTCACTCCTTTT 5404

1050 GluLeuLeuTyrGlyGlyProProProLeuAlaGlu.........IleAl 1063
     |||:::|||||||||||||||||||||||||   :::
5405 GAAGTTCTGTATGGGGGACCTCCCCCTTTAATAAAAGATGGTGGAACATT 5454

1063 aPheAlaHisSerAlaAspValLeuLeuSerGlnProLeuPheSerArgL 1080
               |||:::::::||||||   |||:::   |||:::   :::|
5455 GGTTCCCGATTCAGGCTCTGTCTTACCCTCCTCT...TTGCTTATTCATT 5501

1080 euLysAlaLeuGluTrpValArgGlnArgAlaTrpLysGlnLeuArgGlu 1096
     ||||||||||:::   :::|||   :::   |||   ||||||:::
5502 TAAAGGCCCTGAAAGTAATTAGGACCCAAATTTGGGACCAGCTAAAGACG 5551

1097 AlaTyrSerGlyGlyAspLeuGlnValProHisArgPheGlnValGlyAs 1113
     ||||||:::   |||        ||||||||||:::|||||||||||||||
5552 GCCTACACCCCAGGGACCACCGCAGTACCCCACGAGTTCCAGGTCGGGGA 5601

1113 pSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpL 1130
     |:::|||   ||||||||||||||:::|||:::||||||   |||||||
5602 CCAAGTCTTGGTCAGACGACATCGAACCGGTAGCCTTGAACCACGGTGGA 5651

1130 ysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGlu 1146
     ||||||||||||||||||||||||||||||||||||||||||||||||:::
5652 AGGGACCCTATTTAGTGTTACTAACAACTCCTACGGCAGTGAAAGTTGAC 5701

1147 GlyIleProThrTrpIleHisAlaSerHisValLysProAlaProProPr 1163
     ||||||   :::|||||||||||||||||||||||||   ||||||
5702 GGGATTGCCTCCTGGATCCACGCTTCACACGTCAAGAGGGCCCCCAGTCA 5751

1163 oAsp..............SerGlyTrpArgAlaGluLysThrGluAsnP 1175
      |||               ::::::|||   :::|||   |||:::||||
5752 AGATGAAGAAACCCACGAAGACAACTGGGCTGTGGAAGCCACTGATAACC 5801

1175 roLeuLysLeuArgLeuHisArgLeuValProTyrSerAsnAsnAsnSer 1191
     ||||||||||||||||:::|||      |||            :::
5802 CTCTTAAGCTTCGCTTGCGTCGCAGGAGCCCCCTC.........CATCAC 5842

1192 ProGly 1193
     ||||||
5843 CCTGGA 5848
```

```
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-1

seq_documentation_block:
; Sequence 1, Application US/08716351A
; Patent No. 6033905
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Gibbon Ape Leukemia Virus-Based
;    TITLE OF INVENTION:  Retroviral Vectors
;    NUMBER OF SEQUENCES:  5
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/716,351A
;      FILING DATE:
```

```
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/03784
;      FILING DATE:  06-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian,, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  15280-128-1PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..8535
;      OTHER INFORMATION:  /standard_name= "GaLV SEATO Genome"
US-08-716-351A-1


alignment_scores:
            Quality: 4319.00              Length:   1185
              Ratio:  4.238                Gaps:      6
 Percent Similarity:  85.992   Percent Identity: 68.523

alignment_block:
 US-09-171-553B-5 x US-08-716-351A-1   ..

 Align seg 1/1  to: US-08-716-351A-1  from: 1  to: 8535

       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         |||   :::||||||||||||||||||||||||||||||||||||||||   ||
    2564 GGGAGTCAGGGTTCGGACCCCCTCCCCGAACCTAGGGTAACACTGACTGT 2613

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerV 34
         |||||||   |||:::|||||||||||||||||||||||||||||||||||
    2614 GGAGGGGACCCCCATTGAGTTCCTGGTCGACACCGGAGCTGAACATTCAG 2663

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
         |||||   ||||||:::|||||||:::   :::::::::::::   |||
    2664 TATTGACCCAACCCATGGGAAAAGTAGGGTCCAGACGGACGGTCGTGGAA 2713

      51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
         ||||||||||||:::::   ||||||||||||||:::|||   :::
    2714 GGAGCGACAGGCAGCAAGGTCTACCCCTGGACCACAAAAAGACTTTTAAA 2763

      67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
          :::|||      :::|||||||||||||||||||||||||||||||||||
    2764 AATTGGACATAAACAAGTGACCCACTCCTTCCTGGTCATACCCGAGTGCC 2813

      84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
         ||:::|||||||||||||||||||||||||||||||:::   |||||||||
    2814 CTGCTCCTCTGTTGGGCAGGGACCTCCTAACCAAACTAAAGGCCCAGATC 2863

     101 SerPheGluGlnGlyArgProGluValSerValAsnAsnLysProIleTh 117
         :::|||:::         |||:::|||:::   ::::::::::|||
    2864 CAGTTTTCCGCTGAGGGCCCACAGGTAACATGGGGAGAACGCCCTACTAT 2913

     117 rValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnV 134
             |||:::|||:::|||:::::|||||||||||||:::::      |
    2914 GTGCCTGGTCCTAAACCTGGAAGAAGAATACCGACTACATGAAAAGCCAG 2963

     134 alLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGlnAla 150
         ||   |||::::::         ||||||||:::   ||||||   :::
    2964 TA...CCCTCCTCTATCGACCCATCCTGGCTCCAGCTTTTCCCCACTGTA 3010

     151 TrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnVa 167
         ||||||||   ||||||||||||||||||:::||||||||||||   ||
    3011 TGGGCAGAAAGAGCCGGCATGGGACTAGCCAATCAAGTCCCACCAGTGGT 3060

     167 lIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrProL 184
         |:::::|||:::::::::|||:::|||||||:::||||||||||||||||:
    3061 AGTAGAGCTAAGATCAGGTGCCTCACCAGTGGCTGTTCGACAATATCCAA 3110

     184 euSerArgGluAlaArgGluGlyIleTrpProHisValGlnArgLeuIle 200
         ::|||:::||||||||||||||||||   ||||||:::|||:::::::::
    3111 TGAGCAAAGAAGCTCGGGAAGGTATCAGACCCCACATCCAGAAGTTCCTA 3160

     201 GlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLe 217
         :::   |||:::|||||||||   :::|||||||||||||||||||||||||
    3161 GACCTAGGGGTCTTGGTGCCCTGTCGGTCGCCCTGGAATACCCCTCTGCT 3210

     217 uProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeuA 234
         |||||||:::|||||||||||||||||||||||||||||||||||||||||
    3211 ACCTGTAAAAAAGCCAGGGACCAATGACTATCGGCCAGTTCAAGACCTGA 3260

     234 rgGluValAsnLysArgValGlnAspIleHisProThrValProAsnPro 250
         |||||:::|||||||||||||||||||||||||||||||||||||||||||
    3261 GAGAAATTAATAAAAGGGTACAGGATATTCATCCCACAGTCCCAAACCCT 3310

     251 TyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLe 267
         ||||||||||||||:::||||||||||:::   :::|||||:::|||||
    3311 TACAATCTTCTGAGTTCCCTTCCGCCTAGCTATACTTGGTACTCAGTCTT 3360

     267 uAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnP 284
         ||||||||||||||||||||||||||||||||||||||||||:::|||||||
    3361 AGATCTCAAGGATGCCTTTTTCTGCCTCAGGCTACATCCCAACAGCCAGC 3410

     284 roLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGln 300
         ||||||||||||||||||:::|||||||      |||:::|||||||||
    3411 CGCTGTTCGCGTTCGAGTGGAAAGACCCAGAAAAAGGTAACACAGGTCAG 3460

     301 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePh 317
         ||||||||||||||||||||||||||||||||||||||||||||||:::||
    3461 CTGACCTGGACGCGGCTACCACAAGGGTTCAAGAACTCTCCCACTCTCTT 3510

     317 eAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
         ||||||||||||||||||||||||||   ||||||      :::||||
    3511 CGACGAGGCCCTCCACCGAGATTTGGCTCCCTTTAGGGCCCTCAACCCCC 3560

     334 lnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAlaThr 350
         |||||:::||||||||||||||||||||||||||||:::|||:::   |||
    3561 AGGTGGTGTTACTCCAATATGTGGACGACCTCTTGGTGGCCGCCCCCACA 3610

     351 LysGlnAspCysLeuGluGlyThrLysAlaLeuLeuGluLeuSerAs 367
         :::||||||   :::|||||:::   ||||||   |||||||||
    3611 TATGAAGACTGCAAAAAAGGAACACAGAAGCTCTTACAGGAGTTAAGTAA 3660

     367 pLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluV 384
          ||||||||||||:::|||||||||||||||||||:::|||:::|||||||
    3661 GTTGGGGTACCGGGTATCGGCTAAGAAGGCCCAGCTCTGCCAGAGAGAAG 3710

     384 alThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrGlu 400
         |||||||||||||||||   |||:::   |||:::|||||||||||||||
    3711 TCACCTATCTGGGGTACCTACTCAAGGAAGGAAAAAGATGGCTAACCCCA 3760

     401 AlaArgLysLysThrValValGlnIleProAlaProThrThrAlaLysGl 417
         |||||||||   ||||||::::::|||||:::|||||||||   :::|||
    3761 GCCCGAAAGGCTACTGTTATGAAAATCCCTGTTCCTACGACCCCCAGACA 3810

     417 nValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProG 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    3811 GGTCCGTGAATTTCTAGGCACTGCCGGATTCTGCAGGCTCTGGATCCCTG 3860

     434 lyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGly 450
         ||||||||:::|||||||||||||||||||||||||||||||||||||:::
    3861 GGTTTGCTTCCCTGGCTGCACCCTTGTACCCCCTAACAAAAGAGAGCATC 3910

     451 GlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLy 467
         |||   |||:::   |||||||||:::|||||||||   ||||||||
```

3911 CCTTTTATTTGGACTGAGGAACATCAGCAGGCTTTTGACCACATAAAAAA 3960
467 sAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLysProP 484
3961 AGCCTTGCTGTCAGCCCCTGCATTGGCCCTCCCAGACCTCACCAAGCCAT 4010
484 heThrLeuTyrValAspGluArgLysGlyValAlaArgGlyValLeuThr 500
4011 TCACTCTATATATAGATGAGAGAGCCGGCGTGGCCCGGGGAGTGCTCACT 4060
501 GlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLe 517
4061 CAGACTTTAGGACCCTGGCGGCGGCCAGTAGCATATCTATCAAAAAAACT 4110
517 uAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaAlaV 534
4111 GGATCCGGTGGCCAGCGGGTGGCCAACCTGCCTGAAAGCGGTTGCAGCAG 4160
534 alAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIle 550
4161 TAGCACTCCTTCTCAAAGACGCTGATAAGTTAACCTTGGGACAAAATGTG 4210
551 ThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAs 567
4211 ACTGTGATTGCTTCCCATAGCCTCGAAAGCATCGTGCGGCAACCCCCCGA 4260
567 pArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuT 584
4261 CCGGTGGATGACCAATGCCAGAATGACTCATTACCAGAGCCTGCTGTTAA 4310
584 hrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600
4311 ATGAAAGGGTATCGTTTGCGCCCCCTGCTGTCCTAAACCCAGCTACCCTA 4360
601 LeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeuLe 617
4361 CTTCCAGTCGAGTCGGAAGCCACCCCAGTGCACAGGTGCTCAGAAATCCT 4410
617 uIleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThrG 634
4411 CGCCGAAGAAACTGGAACTCGACGAGACCTAGAAGACCAACCATTGCCCG 4460
634 lyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGluGly 650
4461 GG...GTGCCAACCTGGTATACAGACGGTAGCAGTTTCATCACGGAAGGT 4507
651 LysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAl 667
4508 AAACGGAGAGCAGGGGCCCCGATCGTAGATGGCAAGCGGACGGTATGGGC 4557
667 aSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaL 684
4558 TAGCAGCCTGCCAGAAGGTACGTCAGCCCAGAAGGCTGAACTAGTAGCCT 4607
684 euThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700
4608 TGACGCAGGCATTACGCCTGGCCGAAGGAAAAAACATCAACATCTACACG 4657
701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrLy 717
4658 GACAGCAGGTATGCTTTTGCCACTGCTCATATTCATGGGGCAATATATAA 4707
717 sGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluG 734
4708 GCAGAGGGGGCTGCTCACTTCTGCTGGAAAAGATATCAAAAACAAAGAGG 4757
734 luIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIle 750
4758 AAATTTTGGCCCTGCTAGAGGCCATCCATCTCCCTAGGCGGGTCGCCATT 4807
751 IleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAs 767
4808 ATCCACTGTCCTGGCCACCAGAGGGGAAGTAACCCTGTGGCCACTGGGAA 4857
767 nGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeuL 784
4858 CCGGAGGGCCGACGAGGCTGCAAAGCAAGCCGCCCTGTCGACCAGAGTGC 4907
784 euProIleIleGluThrProLysAlaPro...GluProArgArgGlnTyr 799
4908 TGGCAGGAACTACAAAACCTCAAGAGCCAATCGAGCCCGCTCAAGAAAAG 4957
800 ThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluTh 816
4958 ACCAGGCCG..... 4966
816 rProGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisL 833
4967 .....AGGGAGCTCACCCCTGACC 4985
833 ysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 849
4986 GG...GGAAAAGAATTCATTAAGCGGTTACATCAGTTAACTCACTTAGGA 5032
850 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuAr 866
5033 CCAGAAAAGCTTCTCCAACTAGTGAACCGTACCAGCCTCCTCATCCCGAA 5082
866 gLeuProGlyValAlaAspSerValValLysHisCysValProCysGlnL 883
5083 CCTCCAATCTGCAGTTCGCGAAGTCACCAGTCAGTGTCAGGCTTGTGCCA 5132
883 euValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGly 899
5133 TGACTAATGCGGTCACCACCTACAGAGAGACCGGAAAAAGGCAACGAGGA 5182
900 SerHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAl 916
5183 GATCGACCCGGCGTGTACTGGGAGGTAGACTTCACAGAAATAAAGCCTGG 5232
916 aLysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyT 933
5233 TCGGTATGGAAACAAGTATCTGTTAGTATTCATAGATACTTTCTCCGGAT 5282
933 rpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 949
5283 GGGTAGAAGCTTTTCCTACCAAAACTGAAACGGCCCTAATCGTCTGTAAA 5332
950 LysIleLeuGluGluIleProArgPheGlyIleProLysValIleGl 966
5333 AAAATATTAGAAGAAATTCTACCCCGCTTCGGGATCCCTAAGGTACTCGG 5382
966 ySerAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaL 983
5383 GTCAGACAATGGCCCGGCCTTTGTTGCTCAGGTAAGTCAGGGACTGGCCA 5432
983 ysIleLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSer 999
5433 CTCAACTGGGGATAAATTGGAAGTTACATTGTGCGTATAGACCCCAGAGC 5482
1000 SerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLy 1016
5483 TCAGGTCAGGTAGAAAGAATGAACAGAACAATTAAAGAGACCTTGACCAA 5532
1016 sLeuThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheV 1033
5533 ATTAGCCTTAGAGACCGGTGGAAAAGACTGGGTGACCCTCCTTCCCTTAG 5582
1033 alLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1049
5583 CGCTGCTTAGGGCCAGGAATACCCCTGGCCGGTTTGGTTTAACTCCTTAT 5632
1050 GluLeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHi 1066
5633 GAAATTCTCTATGGAGGACCACCCCCCATACTTGAGTCTGGAGAAACTTT 5682

```
    1066 sSerAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaL 1083
    5683 GGGTCCCGATGATAGATTTCTCCCTGTCTTATTTACTCACTTAAAGGCTT 5732

    1083 euGluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSer 1099
    5733 TAGAAATTGTAAGGACCCAAATCTGGGACCAGATCAAAGAGGTGTATAAG 5782

    1100 GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerValTy 1116
    5783 CCTGGTACCGTAACAATCCCTCACCCGTTCCAGGTCGGGGATCAAGTGCT 5832

    1116 rValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProT 1133
    5833 TGTCAGACGCCATCGACCCAGCAGCCTTGAGCCTCGGTGGAAAGGCCCAT 5882

    1133 yrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIlePro 1149
    5883 ACCTGGTGTTGCTGACTACCCCGACCGCGGTAAAAGTCGATGGTATTGCT 5932

    1150 ThrTrpIleHisAlaSerHisValLysProAlaProPro......ProAs 1164
    5933 GCCTGGGTCCATGCTTCTCACCTCAAACCTGCACCACCTTCGGCACCAGA 5982

    1164 pSerGlyTrpArgAlaGluLysThrGluAsnProLeuLysLeuArgLeuH 1181
    5983 TGAGTCCTGGGAGCTGGAAAAGACTGATCATCCTCTTAAGCTGCGTATTC 6032

    1181 isArg 1182
    6033 GGCGG 6037
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-13

seq_documentation_block:

```
; Sequence 13, Application US/08258420
; Patent No. 5710037
;  GENERAL INFORMATION:
;    APPLICANT:  Nienhuis, Arthur W.
;    APPLICANT:  Vanin, Elio F.
;    TITLE OF INVENTION:  No. 5710037el Retroviral Envelope and LTR and Retroviral Vecto
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan,Cecchi, Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch diskette
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,420
;      FILING DATE:  10-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Olstein, Elliot M.
;      REGISTRATION NUMBER:  24,025
;      REFERENCE/DOCKET NUMBER:  271010-208
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8202 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  viral genome
US-08-258-420-13
```

alignment_scores:

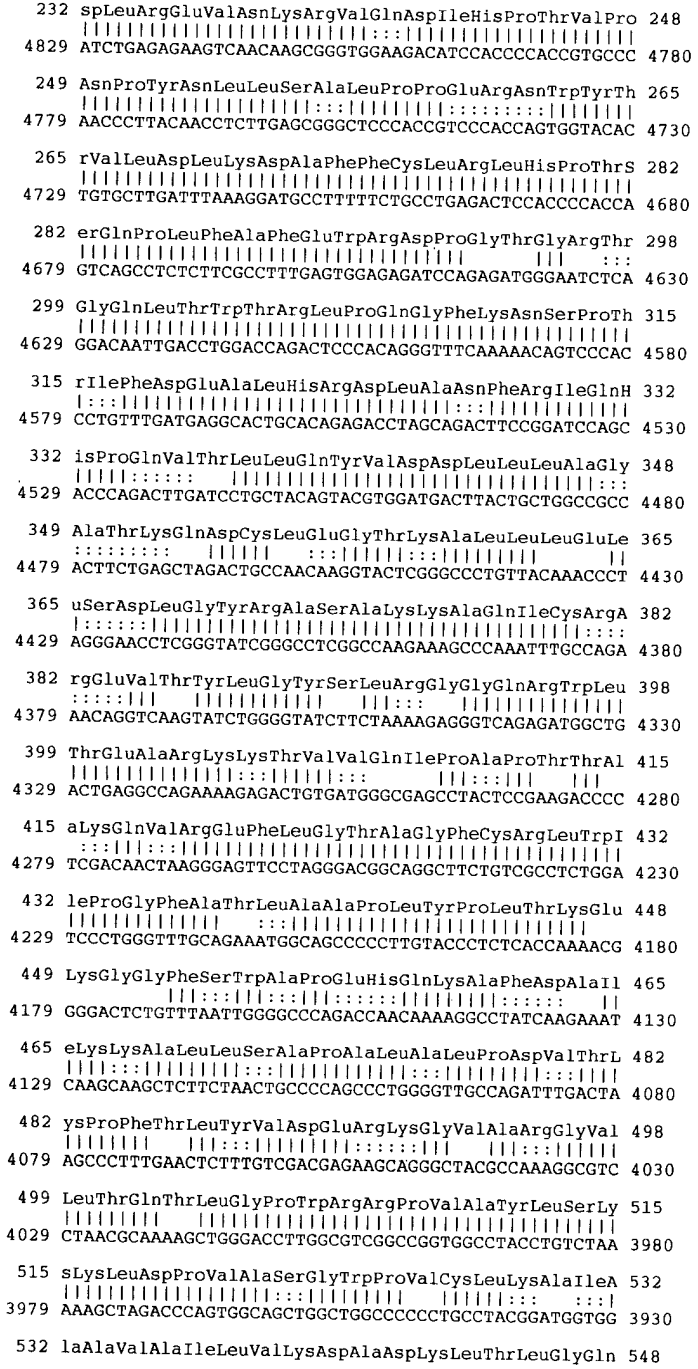
```
           Quality: 3993.50          Length:   1210
             Ratio:  3.982            Gaps:     11
 Percent Similarity:  82.893   Percent Identity: 62.727
```

alignment_block:

US-09-171-553B-5 x US-08-258-420-13/rev ..

Align seg 1/1 to reverse of: US-08-258-420-13 from: 1 to: 8202

```
       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
    5529 GGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGT 5480

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerV 34
    5479 CGGGGGGCAACCCGTCACCTTCCTGGTAGATACTGGGGCCCAACACTCCG 5430

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
    5429 TGCTGACCCAAAATCCTGGACCCCTAAGTGATAAGTCTGCCTGGGTCCAA 5380

      51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
    5379 GGGGCTACTGGAGGAAAGCGGTATCGCTGGACCACGGATCGCAAAGTACA 5330

      67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
    5329 TCTAGCTACCGGTAAGGTCACCCACTCTTTCCTCCATGTACCAGACTGTC 5280

      84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
    5279 CCTATCCTCTGTTAGGAAGAGATTTGCTGACTAAACTAAAAGCCCAAATC 5230

     101 SerPheGlu...GlnGlyArgProGluValSerValAsnAsnLysProIl 116
    5229 CACTTTGAGGGATCAGGAGCTCAGGTTATGGGACCAATGGGGCAGCCCCT 5180

     116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
    5179 GCAAGTGTTGACCCTAAATATAGAAGATGAGTATCGGCTACATGAGACCT 5130

     133 lnValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
    5129 CAAAAGAGCCAGATGTTTCTCTAGGGTCCACATGGCTGTCTGATTTTCCT 5080

     149 GlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPr 165
    5079 CAGGCCTGGGCGGAAACCGGGGGCATGGGACTGGCAGTTCGCCAAGCTCC 5030

     165 oGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnT 182
    5029 TCTGATCATACCTCTGAAAGCAACCTCTACCCCCGTGTCCATAAAACAAT 4980

     182 yrProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArg 198
    4979 ACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGA 4930

     199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPr 215
    4929 CTGTTGGACCAGGGCCTACTGGTACCCTGCCAGTCCCCCTGGAACACGCC 4880

     215 oLeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnA 232
    4879 CCTGCTACCCGTTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGG 4830
```


232 spLeuArgGluValAsnLysArgValGlnAspIleHisProThrValPro 248
4829 ATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCC 4780
249 AsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrTh 265
4779 AACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACAC 4730
265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
4729 TGTGCTTGATTTAAAGGATGCCTTTTTCTGCCTGAGACTCCACCCCACCA 4680
282 erGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
4679 GTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA 4630
299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
4629 GGACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCAC 4580
315 rIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
4579 CCTGTTTGATGAGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGC 4530
332 isProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGly 348
4529 ACCCAGACTTGATCCTGCTACAGTACGTGGATGACTTACTGCTGGCCGCC 4480
349 AlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLe 365
4479 ACTTCTGAGCTAGACTGCCAACAAGGTACTCGGGCCCTGTTACAAACCCT 4430
365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
4429 AGGGAACCTCGGGTATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGA 4380
382 rgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
4379 AACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTG 4330
399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
4329 ACTGAGGCCAGAAAAGAGACTGTGATGGGCGAGCCTACTCCGAAGACCCC 4280
415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
4279 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGA 4230
432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
4229 TCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACG 4180
449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIl 465
4179 GGGACTCTGTTTAATTGGGGCCCAGACCAACAAAAGGCCTATCAAGAAAT 4130
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
4129 CAAGCAAGCTCTTCTAACTGCCCCAGCCCTGGGGTTGCCAGATTTGACTA 4080
482 ysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
4079 AGCCCTTTGAACTCTTTGTCGACGAGAAGCAGGGCTACGCCAAAGGCGTC 4030
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLy 515
4029 CTAACGCAAAAGCTGGGACCTTGGCGTCGGCCGGTGGCCTACCTGTCTAA 3980
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
3979 AAAGCTAGACCCAGTGGCAGCTGGCTGGCCCCCCTGCCTACGGATGGTGG 3930
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
3929 CAGCCATTGCAGTTCTGACAAAAGATGCTGGCAAGCTCACTATGGGACAG 3880
549 AsnIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPr 565
3879 CCGTTGGTCATTCTGGCCCCCCATGCCGTAGAGGCACTAGTTAAGCAACC 3830
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
3829 CCCTGATCGCTGGCTCTCCAATGCCCGGATGACCCATTACCAAGCCCTGC 3780
582 euLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
3779 TCCTGGACACGGACCGGGTCCAGTTCGGGCCAGTAGTGGCCCTAAATCCA 3730
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHi 614
3729 GCTACGCTGCTCCCT...CTGCCTGAGGAGGGGCTGCAACATGACTGCCT 3683
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
3682 TGACATCTTGGCTGAAGCCCACGGAACTAGATCAGATCTTACGGACCAGC 3633
631 roLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrVal 647
3632 CCCTCCCAGACGCCGACCACACCTGGTACACGGATGGGAGCAGCTTCCTG 3583
648 ValGluGlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgTh 664
3582 CAAGAAGGGCAGCGTAAGGCCGGAGCAGCGGTGACCACTGAGACTGAGGT 3533
664 rIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluL 681
3532 AATCTGGGCCAGGGCATTGCCAGCCGGGACATCGGCCCAAAGAGCTGAAC 3483
681 euMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
3482 TGATAGCGCTCACCCAAGCCCTAAAGATGGCAGAAGGTAAGAAGCTAAAT 3433
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
3432 GTTTATACTGATAGCCGTTACGCTTTGCCACCGCCCATATTCATGGAGA 3383
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
3382 AATATACAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAGATCAAGA 3333
731 snLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
3332 ACAAGGACGAGATCTTAGCCCTACTAAAGGCTCTCTTCTTGCCCAAAAGA 3283
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSe 764
3282 CTTAGCATAATTCATTGCCCGGGACATCAAAAAGGAAACAGCGCAGAGGC 3233
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGln.... 779
3232 CAGGGGCAACCGGATGGCCGACCAAGCGGCCCGAGAAGTAGCCACTAGAG 3183
780 .....AlaValAsnLeuLeuProIleIleGluThrProLysAla 792
3182 AAACTCCAGGAACTTCCACACTTCTGATAGAAAACTCAACCCCCTATACC 3133
793 ProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLy 809
3132 CATGAA...CACTTTCACTATACAGTAACTGACACAAAGGATTTGACCAA 3086
809 sIleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSer.... 824
3085 ACTA.....GGAGCCACTTATGACAGTGCGA 3060
825 .....TyrGlyLysGluIleLeuProHisLysGluGly 835

```
3059 AGAAATATTGGGTCACTCAAGGAAAGCCTGTTATGCCTGATCAATTCACC 3010

 836 LeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHi 852
     :::|||   ::::::   :::|||:::|||||||||||||:::   :::
3009 TTTGAGTTACTAGACTTTCTTCACCAATTGACCCACCTCAGCTTCTCAAA 2960

 852 sLeuGlnGlnLeuValArg......ThrSerProTyrHisValLeuArgL 867
         :::   |||::::::        |||||||||:::::|||:::
2959 AACAAAGGCTCTCCTAGAGAGAAGCCCCAGTCCCTACTACATGCTGAACC 2910

 867 euProGlyValAlaAspSerValValLysHisCysValProCysGlnLeu 883
           :::      ::::::::::::   |||      |||
2909 GGGATCGAACACTCAAAAATATCACTGAGACCTGCAAAGCTTGTGCACAA 2860

 884 ValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySe 900
     |||||||||:::   |||   :::      |||   |||:::||||||
2859 GTCAATGCCAGCAAGTCTGCCGTTAAGCAAGGAACTAGGGTCCGCGGGCA 2810

 900 rHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaL 917
      :::||||||:::|||||||||:::|||||||||||||||||||||||:::
2809 TCGGCCTGGCACACACTGGGAGATCGATTTCACCGAGGTAAAACCTGGAT 2760

 917 ysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrp 933
       ||||||   ||||||||||||||||||||||||||||||||||||||||
2759 TGTATGGCTATAAGTATCTTTTAGTTTTTGTAGATACTTTTTCTGGCTGG 2710

 934 ValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLy 950
     :::||||||:::|||||||||||||||||||:::   ||||||:::||||
2709 ATAGAAGCTTTCCCAACTAAGAAAGAAACCGCCAAGGTCGTGACCAAGAA 2660

 950 sIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlyS 967
     |:::|||||||||||||||||||||||||||:::|||:::|||:::|||:
2659 ACTGCTAGAAGAGATCTTCCCTAGGTTCGGCATGCCGCAGGTATTGGGAA 2610

 967 erAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLys 983
     ::|||||||||||||||||||||:::::|||||||||   :::|||
2609 CTGACAATGGGCCTGCCTTCGTCTCCAAGGTGAGTCAGACAGTGGCCGAT 2560

 984 IleLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSe 1000
     :::||||||||||||||||||||||||||||||||||||||||||||||||
2559 CTGTTGGGGATTGATTGGAAATTACATTGTGCATACAGACCCCAAAGCTC 2510

1000 rGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysL 1017
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2509 AGGTCAGGTAGAAAGAATGAATAGGACCATCAAGGAGACTTTAACTAAAT 2460

1017 euThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVal 1033
     |||||      ||||||   :::|||||:::   |||||||||::::::
2459 TAACGCTTGCAACTGGCTCTAGAGACTGGGTGCTCCTACTCCCCTTAGCC 2410

1034 LeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGl 1050
     |||:::|||:::|||||||||||||||      ||||||||||||||||
2409 CTGTACCGAGCCCGCAACACGCCGGGCCCCCATGGCCTCACCCCATATGA 2360

1050 uLeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisS 1067
     |:::|||||||||:::|||||||||||||:::::::::
2359 GATCTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGACCCTGACA 2310

1067 erAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeu 1083
       :::   |||      |||      |||   ::::::|||:::||||||
2309 TGACCAGAGTTACTAACAGCCCCTCTCTCCAAGCTCACTTACAGGCCCTT 2260

1084 GluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGl 1100
     :::   |||:::::::::::::||||||   |||      ||||||:::
2259 CAAGCAGTACAACGAGAGGTCTGGAAGCCACTGGCTGCGGCCTATCAGGA 2210

1100 yGlyAspLeuGln......ValProHisArgPheGlnValGlyAspServ 1115
         :::   |||      :::|||||||   |||:::|||||||||:::|
2209 CCAGCAAGACCAGCCAGTGATACCACACCCCTTCCGTGTCGGCGACACCG 2160

1115 alTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGly 1131
     ||:::|||||||||||||:::::   |||||||||   ||||||||||||
2159 TGTGGGTACGCCGGCACCAGACTAAGAACTTGGAACCTCGTTGGAAAGGA 2110

1132 ProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIl 1148
     ||||||   |||||||||||||||||||||||:::||||||:::||||||
2109 CCCTATACCGTCCTGCTGACCACCCCCACCGCTCTCAAAGTAGACGGCAT 2060

1148 eProThrTrpIleHisAlaSerHisValLysProAlaProProProAspS 1165
     |   :::|||||||||||||:::|||||||||   |||      |||  :
2059 CGCTGCGTGGATCCACGCCGCTCACGTAAAGGCGGCGACAACCCCTCCGG 2010

1165 erGly.................TrpArgAlaGluLysThrGluAsnPro 1175
     ::|||                 |||:::::::::::::::::||||||
2009 CCGGAACAGCATCAGGACCGACATGGAAGGTCCAGCGTTCTCAAAACCCC 1960

1176 LeuLysLeuArgLeuHisArgLeuValPro 1185
     ||||||:::|||||||   |||   :::|||
1959 TTAAAGATAAGATTAACCCGTGGGGCCCCC 1930
```

```
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-8

seq_documentation_block:
; Sequence 8, Application US/08110300A
; Patent No. 5643756
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,300A
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-110-300A-8


alignment_scores:
            Quality: 3980.00              Length:   1206
              Ratio:  3.992                 Gaps:     12
 Percent Similarity:  82.670   Percent Identity: 62.687

alignment_block:
 US-09-171-553B-5 x US-08-110-300A-8   ..

 Align seg 1/1  to: US-08-110-300A-8  from: 1  to: 8323

       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
```

```
 614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
     :::::::|||   |||      |||:::|||   ||||||||||||   |
4083 TGACATCTTGGCTGAAGCCCACGGAACTAGACCAGATCTTACGGACCAGC 4132

 631 roLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrVal 647
     |||||                |||||:::||||||||||||||:::::
4133 CTCTCCCAGACGCTGACCACACCTGGTACACAGATGGGAGCAGCTTCCTG 4182

 648 ValGluGlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgTh 664
       |||||||:::|||   |||||||||||||||:::      |||:::::
4183 CAAGAGGGGCAGCGCAAGGCCGGAGCAGCAGTAACCACCGAGACCGAGGT 4232

 664 rIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluL 681
     ::::||||||:::::||||||   |||||||||||||||:::|||||||
4233 AGTCTGGGCCAAAGCACTGCCAGCCGGGACATCGGCCCAAAGAGCTGAGT 4282

 681 euMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
     ||:::|||||||||||||||||:::::::|||||||||||:::::::|||
4283 TGATAGCGCTCACCCAAGCCTTAAAAATGGCAGAAGGTAAGAAGCTGAAT 4332

 698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
     :::|||||||||||||||||||||||||||||||||||:::||||||
4333 GTTTACACCGATAGCCGTTATGCTTTTGCCACTGCCCATATTCACGGAGA 4382

 714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
      ||||||:::::|||||||||||||||||   |||:::||||||||||
4383 AATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAAATCAAAA 4432

 731 snLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
     |||||:::|||||||||:::|||||||:::||||||   ||||||||||||
4433 ATAAGGACGAGATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGA 4482

 748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSe 764
     |||:::|||||||||||||||||||||||||||:::::::        ::
4483 CTTAGCATAATTCATTGCCCGGGACATCAGAAGGGAAACCGCGCGGAGGC 4532

 764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaV 781
     :||||||||||:::|||||||||:::::::|||:::::::::|||
4533 AAGGGGCAACAGGATGGCCGACCAAGCGGCCCGAGAAGTAGCCACTAGA. 4581

 781 alAsnLeuLeuProIleIleGluThrProLys..................  791
                          |||||||||:::
4582 ....................GAAACTCCAGAGACTTCCACACTTCTGATA 4611

 792 .........AlaProGluProArgArg......GlnTyrThrLeuGluAs 803
             ||||||      ::::::      :::|||||:::   ||
4612 GAAAATTCAGCCCCCTATACTCATGAACATTTTCACTATACGGTGACTGA 4661

 803 pTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGlyT 820
     |   :::::::::   |||:::          :::   :::
4662 CATAAAAGATCTGACTAAACTAGGGGCCACTTATGACGATGCAAAG...A 4708

 820 hrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeu 836
       |||::::::      ||||||   :::::::|||   :::      :::
4709 AGTGTTGGGTTTATCAGGGAAAGCCTGTAATGCCTGATCAATTCACCTTT 4758

 837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLe 853
     |||   ::::::   :::|||:::|||||||||||||:::   :::
4759 GAACTATTAGATTTTCTTCATCAATTGACCCACCTCAGTTTCTCAAAAAC 4808

 853 uGlnGlnLeuValArgThrSer......ProTyrHisValLeuArgLeuP 868
      :::   |||::::::   :::      ||||||::::::|||:::
4809 AAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTACATGCTGAACCGGG 4858

 868 roGlyValAlaAspSerValValLysHisCysValProCysGlnLeuVal 884
         :::      ::::::::::::   |||      |||      |||
4859 ATCGAACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAGGTC 4908

 885 AsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySerHi 901
     ||||||:::   |||   :::      |||   |||:::||||||   ::
4909 AATGCCAGCAAGTCTGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCG 4958
```

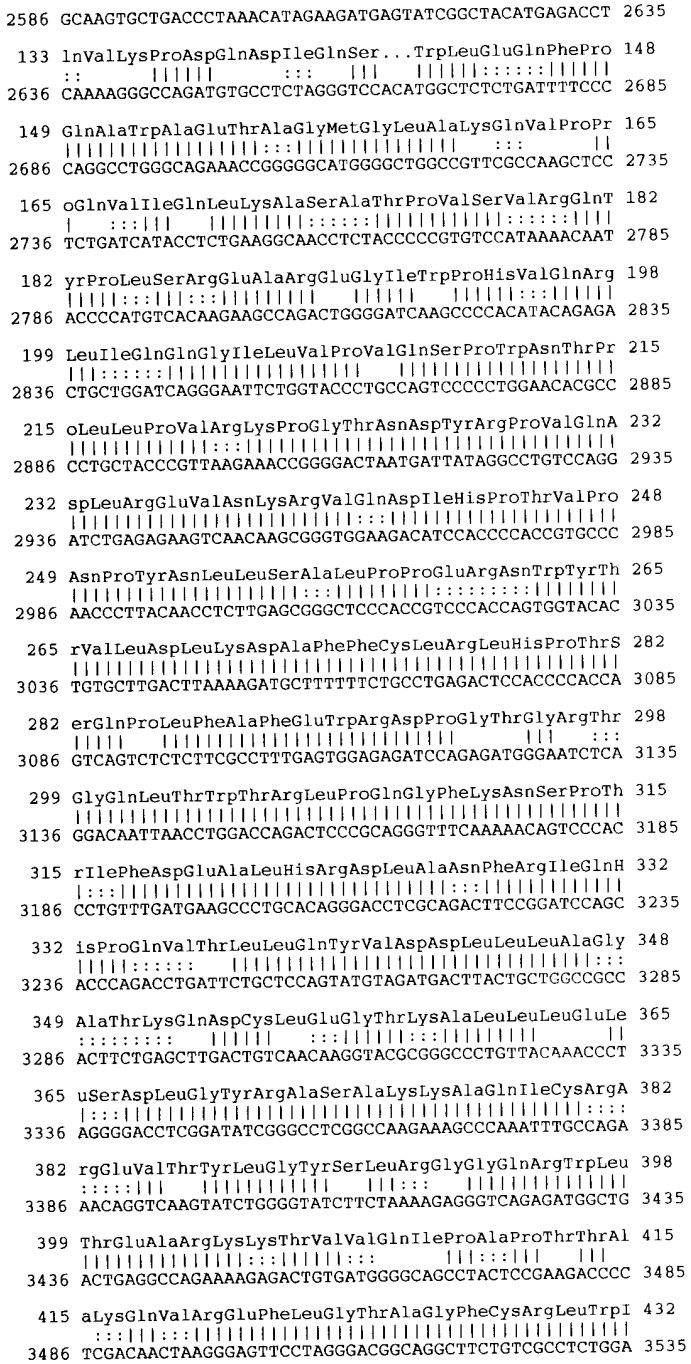
```
 901 sProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLysT 918
     :||||||:::||||||||||:::|||||||||||||||||||||:::   |
4959 ACCCGGCACCCACTGGGAAATTGATTTCACTGAGGTAAAACCTGGCCTGT 5008

 918 yrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpVal 934
     |||||   |||||||||||||||||:::|||||||||||||||||||||
5009 ATGGGTATAAATATCTTTTAGTTTTCATAGACACTTTCTCTGGATGGGTA 5058

 935 GluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLysIl 951
     ||||||:::|||||||||||||||||||:::   ||||||:::||||||::
5059 GAAGCTTTCCCAACCAAGAAAGAAACTGCCAAAGTTGTAACCAAGAAGCT 5108

 951 eLeuGluGluIleProProArgPheGlyIleProLysValIleGlySerA 968
     :|||||||||||||||||||||||||||||:::|||:::|||:::|||:::|
5109 ACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAGGTATTGGGAACCG 5158

 968 spAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIle 984
     ||||||||||||||||||||:::::|||||||||   :::|||   :::
5159 ACAATGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGATTTA 5208

 985 LeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerGl 1001
     ||||||:::|||||||||||||||||||||||||||||||||||||||||
5209 CTGGGGGTTGATTGGAAACTACATTGTGCTTACAGACCCCAGAGTTCAGG 5258

1001 yGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuT 1018
     ||||||||||||||||||||||||||||||||||||||||||||||||||
5259 TCAGGTAGAAAGAATGAATAGGACAATCAAGGAGACTTTAACTAAATTGA 5308

1018 hrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeu 1034
     ||      ||||||   :::|||||:::   |||||||||::::::|||
5309 CGCTTGCAACTGGCTCTAGGGACTGGGTGCTCCTGCTTCCCCTAGCCCTG 5358

1035 PheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGluLe 1051
     :::|||:::|||||||||||||||      ||||||||||||||||||::
5359 TATCGAGCCCGCAACACGCCGGGCCCCCATGGTCTCACCCCATATGAAAT 5408

1051 uLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisSerA 1068
     :|||||||||:::|||||||||||:::::::::                |
5409 CTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGATCCTGACATGG 5458

1068 laAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGlu 1084
     ||   |||      :::      |||   :::::::|||:::|||||||
5459 CAAAGGTTACTCATAACCCCTCTCTCCAAGCCCATTTACAGGCACTCTAC 5508

1085 TrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGlyGl 1101
       |||:::::::::::::|||:::   |||      ||||||:::
5509 CTGGTCCAGCACGAAGTCTGGAGACCGTTGGCGGCAGCTTACCAA...GA 5555

1101 yAspLeuGln.........ValProHisArgPheGlnValGlyAspServ 1115
      :::|||:::         |||||||||   |||:::||||||||||:::|
5556 ACAACTGGACCGGCCGGTAGTGCCTCACCCTTTCCGAGTCGGTGACACAG 5605

1115 alTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGly 1131
     ||:::|||||||||||||:::::   |||||||||   ||||||||||||
5606 TGTGGGTCCGCAGACACCAAACTAAAAATCTAGAACCCCGCTGGAAAGGA 5655

1132 ProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIl 1148
     ||||||   |||||||||||||||||||||||||:::|||||||:::|||||
5656 CCTTATACCGTCCTACTGACTACCCCCACCGCTCTCAAAGTGGACGGCAT 5705

1148 eProThrTrpIleHisAlaSerHisValLysProAla.............  1160
     |   :::|||||||||||||:::|||||||||   |||
5706 TGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGACACCAGGATTG 5755

1161 ..ProProProAspSerGlyTrpArgAlaGluLysThrGluAsnProLeu 1176
       ||||||   ::|||   ||||||:::::::::::::::|||||||||
5756 AGCCACCATCGGAATCGACATGGCGTGTTCAACGCTCTCAAAATCCCCTA 5805

1177 LysLeuArgLeuHisArg 1182
     |||:::||||||   |||
5806 AAGATAAGATTGACCCGC 5823
```

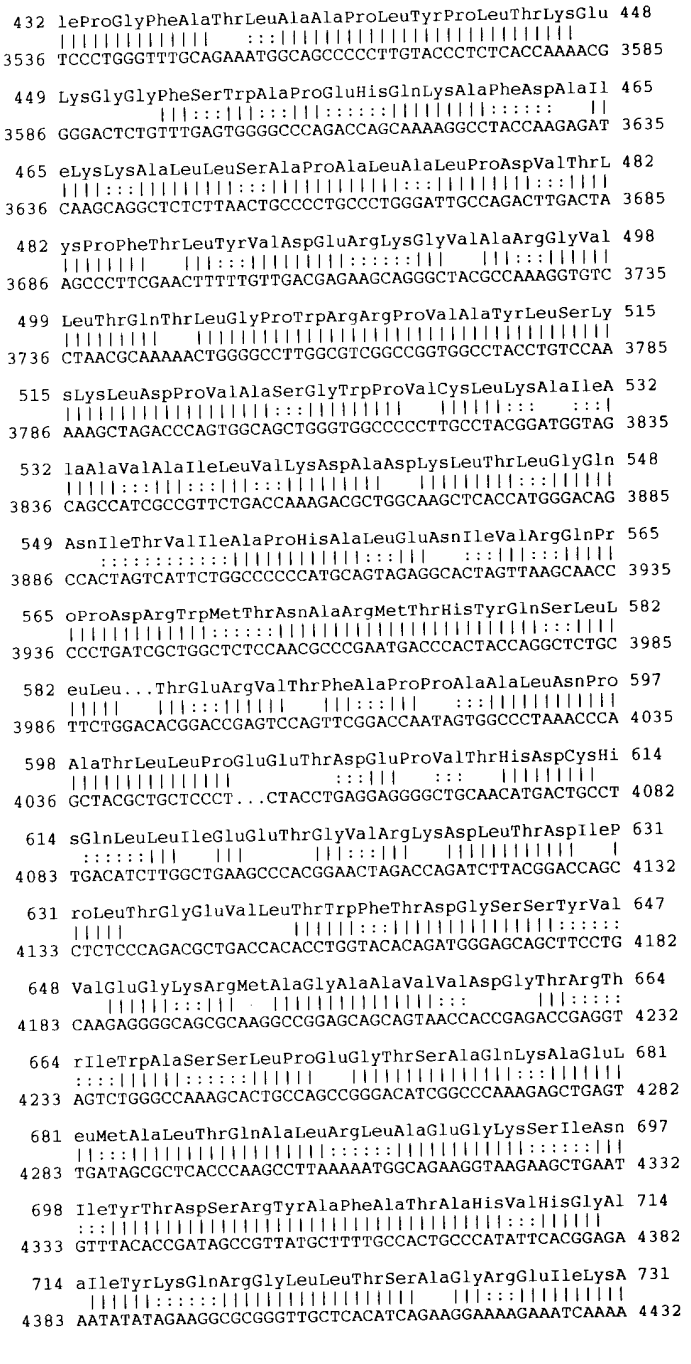
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-8

```
seq_documentation_block:
; Sequence 8, Application US/08886642
; Patent No. 5952474
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  45 Rockefeller Plaza, Suite 2800
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/886,642
;      FILING DATE:  01-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/110,300
;      FILING DATE:  20-AUG-1993
;      APPLICATION NUMBER:  07/938,100
;      FILING DATE:  28-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hone, William J.
;      REGISTRATION NUMBER:  26,739
;      REFERENCE/DOCKET NUMBER:  07763/010002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212/765-5070
;      TELEFAX:  212/258-2291
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-886-642-8


alignment_scores:
            Quality: 3980.00              Length:   1206
              Ratio:  3.992                Gaps:     12
 Percent Similarity:  82.670   Percent Identity: 62.687

alignment_block:
 US-09-171-553B-5 x US-08-886-642-8   ..

 Align seg 1/1  to: US-08-886-642-8  from: 1  to: 8323

       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         |||   :::|||:::::::||   |||||||||||||:::|||||||:::||
    2236 GGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAGAGT 2285

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisServ 34
         |   ||||||||||||   |||||||||||||||||||||:::|||||||
    2286 CGGGGGGCAACCCGTCACCTTCCTAGTGGATACTGGGGCCCAACACTCCG 2335

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
         |||||   |||       |||   |||:::::::|||:::::::|||||:::
    2336 TGCTGACCCAAAATCCTGGACCCCTAAGTGACAAGTCTGCCTGGGTCCAA 2385

      51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
         ||||||||||||   :::::::|||   |||||||||   |||   |||
    2386 GGGGCTACTGGAGGGAAGCGGTATCGCTGGACCACGGATCGCCGAGTGCA 2435

      67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84

         |||:::::::|||:::|||||||||||||||||||   :::|||:::||||
    2436 CCTAGCCACCGGTAAGGTCACCCATTCTTTCCTCCATGTACCAGATTGCC 2485

      84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
         ||   |||||||||||||||||||||||||||||:::   |||||||||
    2486 CCTATCCTCTGCTAGGAAGAGATTTGCTGACTAAACTAAAAGCCCAAATT 2535

     101 SerPheGlu...GlnGlyArgProGluValSerValAsnAsnLysProIl 116
         ||||||   :::|||       |||:::      ::::::|||::
    2536 CACTTTGAGGGATCAGGAGCTCAGGTTGTGGGACCAATGGGACAGCCCCT 2585

     116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
         :   ||||||||||||:::::::::|||||||||||||||||:::::   :
    2586 GCAAGTGCTGACCCTAAACATAGAAGATGAGTATCGGCTACATGAGACCT 2635

     133 lnValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
         ::      ||||||      :::   |||   ||||||:::::||||||
    2636 CAAAAGGGCCAGATGTGCCTCTAGGGTCCACATGGCTCTCTGATTTTCCC 2685

     149 GlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPr 165
         ||||||||||||||||:::|||||||||||||||   :::      ||
    2686 CAGGCCTGGGCAGAAACCGGGGGCATGGGGCTGGCCGTTCGCCAAGCTCC 2735

     165 oGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnT 182
         |   :::|||   |||||||||:::::|||||||||||:::::::||||
    2736 TCTGATCATACCTCTGAAGGCAACCTCTACCCCCGTGTCCATAAAACAAT 2785

     182 yrProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArg 198
         |||||:::|||:::||||||||||   ||||||   ||||||:::||||||
    2786 ACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGA 2835

     199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPr 215
         |||:::::|||||||||||||||||||   ||||||||||||||||||||
    2836 CTGCTGGATCAGGGAATTCTGGTACCCTGCCAGTCCCCCTGGAACACGCC 2885

     215 oLeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnA 232
         ||||||||||||:::|||||||||||||||||||||||||||||||||||
    2886 CCTGCTACCCGTTAAGAAACCGGGGACTAATGATTATAGGCCTGTCCAGG 2935

     232 spLeuArgGluValAsnLysArgValGlnAspIleHisProThrValPro 248
         ||||||||||||||||||||||||||||:::||||||||||||||||||||
    2936 ATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCC 2985

     249 AsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrTh 265
         ||||||||||||||||||||||:::|||||||||:::::::::|||||||
    2986 AACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACAC 3035

     265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3036 TGTGCTTGACTTAAAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCA 3085

     282 erGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
         |||||   ||||||||||||||||||||||||||      |||   :::
    3086 GTCAGTCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA 3135

     299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3136 GGACAATTAACCTGGACCAGACTCCCGCAGGGTTTCAAAAACAGTCCCAC 3185

     315 rIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
         |:::|||||||||||||||||||||||||||||:::|||||||||||||||
    3186 CCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGGATCCAGC 3235

     332 isProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGly 348
         |||||::::::   ||||||||||||||||||||||||||||||||:::
    3236 ACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCC 3285

     349 AlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLe 365
         :::::::::   ||||||   :::|||||||:::|||||||||      ||
    3286 ACTTCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCT 3335

     365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
         |:::|||||||||||||||||||||||||||||||||||||||||||::::
```

3336 AGGGGACCTCGGATATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGA 3385
382 rgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
3386 AACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTG 3435
399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
3436 ACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 3485
415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
3486 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGA 3535
432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
3536 TCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACG 3585
449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIl 465
3586 GGGACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGAT 3635
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
3636 CAAGCAGGCTCTCTTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTA 3685
482 ysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
3686 AGCCCTTCGAACTTTTTGTTGACGAGAAGCAGGGCTACGCCAAAGGTGTC 3735
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLy 515
3736 CTAACGCAAAAACTGGGGCCTTGGCGTCGGCCGGTGGCCTACCTGTCCAA 3785
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
3786 AAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAG 3835
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
3836 CAGCCATCGCCGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAG 3885
549 AsnIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPr 565
3886 CCACTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTTAAGCAACC 3935
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
3936 CCCTGATCGCTGGCTCTCCAACGCCCGAATGACCCACTACCAGGCTCTGC 3985
582 euLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
3986 TTCTGGACACGGACCGAGTCCAGTTCGGACCAATAGTGGCCCTAAACCCA 4035
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHi 614
4036 GCTACGCTGCTCCCT...CTACCTGAGGAGGGGCTGCAACATGACTGCCT 4082
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
4083 TGACATCTTGGCTGAAGCCCACGGAACTAGACCAGATCTTACGGACCAGC 4132
631 roLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrVal 647
4133 CTCTCCCAGACGCTGACCACACCTGGTACACAGATGGGAGCAGCTTCCTG 4182
648 ValGluGlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgTh 664
4183 CAAGAGGGGCAGCGCAAGGCCGGAGCAGCAGTAACCACCGAGACCGAGGT 4232
664 rIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluL 681
4233 AGTCTGGGCCAAAGCACTGCCAGCCGGGACATCGGCCCAAAGAGCTGAGT 4282
681 euMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
4283 TGATAGCGCTCACCCAAGCCTTAAAAATGGCAGAAGGTAAGAAGCTGAAT 4332
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
4333 GTTTACACCGATAGCCGTTATGCTTTTGCCACTGCCCATATTCACGGAGA 4382
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
4383 AATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAAATCAAAA 4432
731 snLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
4433 ATAAGGACGAGATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGA 4482
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSe 764
4483 CTTAGCATAATTCATTGCCCGGGACATCAGAAGGGAAACCGCGCGGAGGC 4532
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaV 781
4533 AAGGGGCAACAGGATGGCCGACCAAGCGGCCCGAGAAGTAGCCACTAGA. 4581
781 alAsnLeuLeuProIleIleGluThrProLys..... 791
4582 .....GAAACTCCAGAGACTTCCACACTTCTGATA 4611
792 .....AlaProGluProArgArg.....GlnTyrThrLeuGluAs 803
4612 GAAAATTCAGCCCCCTATACTCATGAACATTTTCACTATACGGTGACTGA 4661
803 pTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGlyT 820
4662 CATAAAAGATCTGACTAAACTAGGGGCCACTTATGACGATGCAAAG...A 4708
820 hrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeu 836
4709 AGTGTTGGGTTTATCAGGGAAAGCCTGTAATGCCTGATCAATTCACCTTT 4758
837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLe 853
4759 GAACTATTAGATTTTCTTCATCAATTGACCCACCTCAGTTTCTCAAAAAC 4808
853 uGlnGlnLeuValArgThrSer.....ProTyrHisValLeuArgLeuP 868
4809 AAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTACATGCTGAACCGGG 4858
868 roGlyValAlaAspSerValValLysHisCysValProCysGlnLeuVal 884
4859 ATCGAACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAGGTC 4908
885 AsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySerHi 901
4909 AATGCCAGCAAGTCTGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCG 4958
901 sProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLysT 918
4959 ACCCGGCACCCACTGGGAAATTGATTTCACTGAGGTAAAACCTGGCCTGT 5008
918 yrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpVal 934
5009 ATGGGTATAAATATCTTTTAGTTTTCATAGACACTTTCTCTGGATGGGTA 5058
935 GluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLysIl 951
5059 GAAGCTTTCCCAACCAAGAAAGAAACTGCCAAAGTTGTAACCAAGAAGCT 5108
951 eLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerA 968
5109 ACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAGGTATTGGGAACCG 5158

```
     968 spAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIle 984
         ||||||||||||||||||||:::::|||||||||   :::|||   :::
    5159 ACAATGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGATTTA 5208

     985 LeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerGl 1001
         ||||||:::|||||||||||||||||||||||||||||||||||||||||
    5209 CTGGGGGTTGATTGGAAACTACATTGTGCTTACAGACCCCAGAGTTCAGG 5258

    1001 yGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuT 1018
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    5259 TCAGGTAGAAAGAATGAATAGGACAATCAAGGAGACTTTAACTAAATTGA 5308

    1018 hrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeu 1034
         ||      ||||||   :::|||||||:::   |||||||||:::::|||
    5309 CGCTTGCAACTGGCTCTAGGGACTGGGTGCTCCTGCTTCCCCTAGCCCTG 5358

    1035 PheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGluLe 1051
         :::|||:::|||||||||||||||      ||||||||||||||||||::
    5359 TATCGAGCCCGCAACACGCCGGGCCCCCATGGTCTCACCCCATATGAAAT 5408

    1051 uLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisSerA 1068
         :|||||||||:::|||||||||||||:::::::::                |
    5409 CTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGATCCTGACATGG 5458

    1068 laAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGlu 1084
         ||   |||     :::      |||   ::::::|||:::||||||
    5459 CAAAGGTTACTCATAACCCCTCTCTCCAAGCCCATTTACAGGCACTCTAC 5508

    1085 TrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGlyGl 1101
            |||:::::::::::::|||:::   |||      ||||||:::
    5509 CTGGTCCAGCACGAAGTCTGGAGACCGTTGGCGGCAGCTTACCAA...GA 5555

    1101 yAspLeuGln.........ValProHisArgPheGlnValGlyAspSerV 1115
          :::|||:::        |||||||||   |||:::||||||||:::|
    5556 ACAACTGGACCGGCCGGTAGTGCCTCACCCTTTCCGAGTCGGTGACACAG 5605

    1115 alTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGly 1131
         ||:::|||||||||||||:::::   |||||||||   ||||||||||||
    5606 TGTGGGTCCGCAGACACCAAACTAAAAATCTAGAACCCCGCTGGAAAGGA 5655

    1132 ProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIl 1148
         ||||||   ||||||||||||||||||||||:::|||||:::||||||
    5656 CCTTATACCGTCCTACTGACTACCCCCACCGCTCTCAAAGTGGACGGCAT 5705

    1148 eProThrTrpIleHisAlaSerHisValLysProAla..... 1160
         |   :::|||||||||||||:::||||||||   |||
    5706 TGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGACACCAGGATTG 5755

    1161 ..ProProProAspSerGlyTrpArgAlaGluLysThrGluAsnProLeu 1176
           ||||||   :::|||   ||||||:::::::::::::::::|||||||
    5756 AGCCACCATCGGAATCGACATGGCGTGTTCAACGCTCTCAAAATCCCCTA 5805

    1177 LysLeuArgLeuHisArg 1182
         |||:::||||||   |||
    5806 AAGATAAGATTGACCCGC 5823

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-8

seq_documentation_block:
; Sequence 8, Application PC/TUS9308041
;  GENERAL INFORMATION:
;    APPLICANT:  The Public Health Research Institute of the City of
;    APPLICANT:  New York, Inc.
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08041
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/938,100
;      FILING DATE:  28-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/_____
;      FILING DATE:  20-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-08041-8


alignment_scores:
            Quality: 3980.00              Length:   1206
              Ratio:  3.992                Gaps:     12
 Percent Similarity:  82.670   Percent Identity: 62.687

alignment_block:
 US-09-171-553B-5 x PCT-US93-08041-8   ..

 Align seg 1/1  to: PCT-US93-08041-8  from: 1  to: 8323

       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         |||   :::|||:::::::|||   |||||||||||:::|||||:::||
    2236 GGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAGAGT 2285

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisServ 34
         |   ||||||||||||   |||||||||||||||||||||:::|||||||
    2286 CGGGGGGCAACCCGTCACCTTCCTAGTGGATACTGGGGCCCAACACTCCG 2335

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
         |||||   |||       |||   |||:::::|||:::::|||||:::
    2336 TGCTGACCCAAAATCCTGGACCCCTAAGTGACAAGTCTGCCTGGGTCCAA 2385

      51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
         ||||||||||||   :::::::|||   |||||||||   |||   |||
    2386 GGGGCTACTGGAGGGAAGCGGTATCGCTGGACCACGGATCGCCGAGTGCA 2435

      67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
         |||:::::::|||:::||||||||||||||||||   :::|||:::||||
    2436 CCTAGCCACCGGTAAGGTCACCCATTCTTTCCTCCATGTACCAGATTGCC 2485

      84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
         ||   |||||||||||||||||||||||||||||:::   |||||||||
    2486 CCTATCCTCTGCTAGGAAGAGATTTGCTGACTAAACTAAAAGCCCAAATT 2535

     101 SerPheGlu...GlnGlyArgProGluValSerValAsnAsnLysProIl 116
           ||||||   :::|||         |||:::      :::::::|||::
    2536 CACTTTGAGGGATCAGGAGCTCAGGTTGTGGGACCAATGGGACAGCCCCT 2585

     116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
         :   ||||||||||||:::::::::|||||||||||||||||:::::   :
```


2586 GCAAGTGCTGACCCTAAACATAGAAGATGAGTATCGGCTACATGAGACCT 2635
133 lnValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
2636 CAAAAGGGCCAGATGTGCCTCTAGGGTCCACATGGCTCTCTGATTTTCCC 2685
149 GlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPr 165
2686 CAGGCCTGGGCAGAAACCGGGGGCATGGGGCTGGCCGTTCGCCAAGCTCC 2735
165 oGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnT 182
2736 TCTGATCATACCTCTGAAGGCAACCTCTACCCCCGTGTCCATAAAACAAT 2785
182 yrProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArg 198
2786 ACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGA 2835
199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPr 215
2836 CTGCTGGATCAGGGAATTCTGGTACCCTGCCAGTCCCCCTGGAACACGCC 2885
215 oLeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnA 232
2886 CCTGCTACCCGTTAAGAAACCGGGGACTAATGATTATAGGCCTGTCCAGG 2935
232 spLeuArgGluValAsnLysArgValGlnAspIleHisProThrValPro 248
2936 ATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCC 2985
249 AsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrTh 265
2986 AACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACAC 3035
265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
3036 TGTGCTTGACTTAAAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCA 3085
282 erGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
3086 GTCAGTCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA 3135
299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
3136 GGACAATTAACCTGGACCAGACTCCCGCAGGGTTTCAAAAACAGTCCCAC 3185
315 rIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
3186 CCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGGATCCAGC 3235
332 isProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGly 348
3236 ACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCC 3285
349 AlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLe 365
3286 ACTTCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCT 3335
365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
3336 AGGGGACCTCGGATATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGA 3385
382 rgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
3386 AACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTG 3435
399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
3436 ACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 3485
415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
3486 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGA 3535

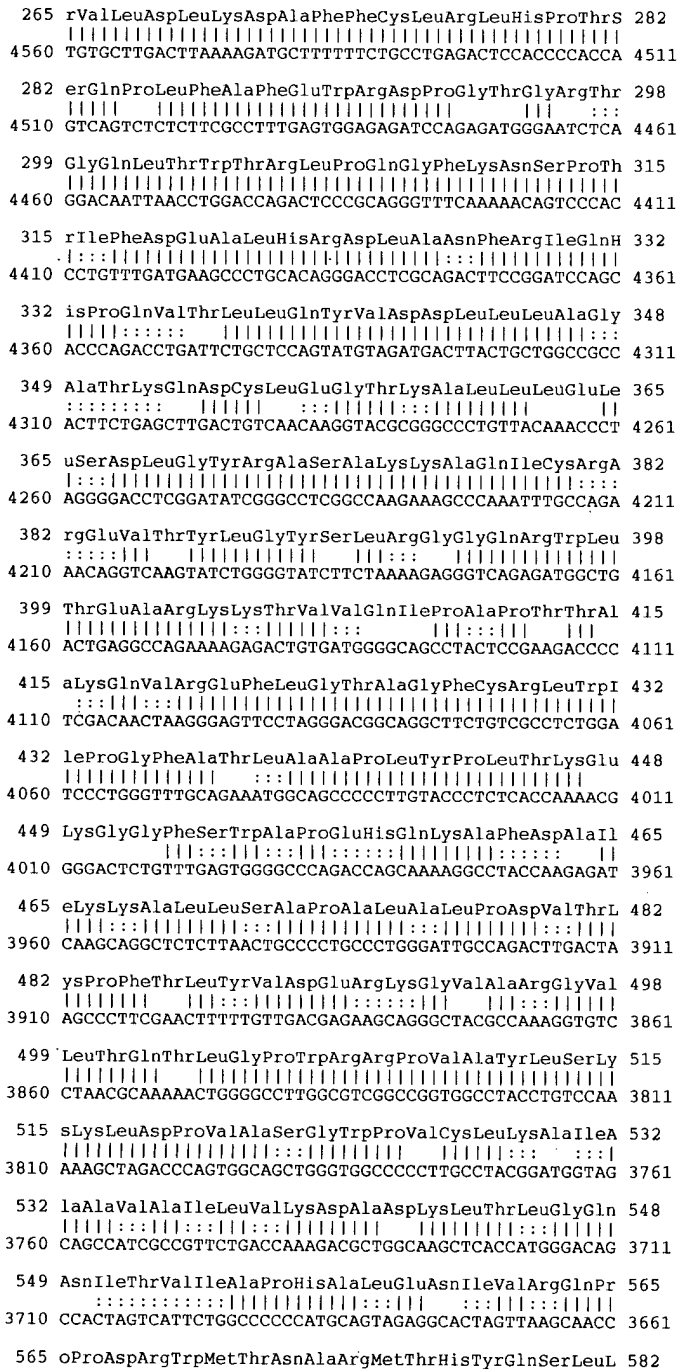

432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
3536 TCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACG 3585
449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIl 465
3586 GGGACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGAT 3635
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
3636 CAAGCAGGCTCTCTTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTA 3685
482 ysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
3686 AGCCCTTCGAACTTTTTGTTGACGAGAAGCAGGGCTACGCCAAAGGTGTC 3735
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLy 515
3736 CTAACGCAAAAACTGGGGCCTTGGCGTCGGCCGGTGGCCTACCTGTCCAA 3785
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
3786 AAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAG 3835
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
3836 CAGCCATCGCCGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAG 3885
549 AsnIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPr 565
3886 CCACTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTTAAGCAACC 3935
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
3936 CCCTGATCGCTGGCTCTCCAACGCCCGAATGACCCACTACCAGGCTCTGC 3985
582 euLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
3986 TTCTGGACACGGACCGAGTCCAGTTCGGACCAATAGTGGCCCTAAACCCA 4035
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHi 614
4036 GCTACGCTGCTCCCT...CTACCTGAGGAGGGGCTGCAACATGACTGCCT 4082
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
4083 TGACATCTTGGCTGAAGCCCACGGAACTAGACCAGATCTTACGGACCAGC 4132
631 roLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrVal 647
4133 CTCTCCCAGACGCTGACCACACCTGGTACACAGATGGGAGCAGCTTCCTG 4182
648 ValGluGlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgTh 664
4183 CAAGAGGGGCAGCGCAAGGCCGGAGCAGCAGTAACCACCGAGACCGAGGT 4232
664 rIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluL 681
4233 AGTCTGGGCCAAAGCACTGCCAGCCGGGACATCGGCCCAAAGAGCTGAGT 4282
681 euMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
4283 TGATAGCGCTCACCCAAGCCTTAAAAATGGCAGAAGGTAAGAAGCTGAAT 4332
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
4333 GTTTACACCGATAGCCGTTATGCTTTTGCCACTGCCCATATTCACGGAGA 4382
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
4383 AATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAAATCAAAA 4432

```
 731 snLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
     |||||:::||||||||||:::||||||:::||||||   ||||||||||||
4433 ATAAGGACGAGATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGA 4482

 748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSe 764
     |||:::|||||||||||||||||||||||||||||:::::::      ::
4483 CTTAGCATAATTCATTGCCCGGGACATCAGAAGGGAAACCGCGCGGAGGC 4532

 764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaV 781
     :|||||||||:::|||||||||:::::|||:::::::::|||
4533 AAGGGGCAACAGGATGGCCGACCAAGCGGCCCGAGAAGTAGCCACTAGA. 4581

 781 alAsnLeuLeuProIleIleGluThrProLys..................  791
                          ||||||||||:::
4582 ....................GAAACTCCAGAGACTTCCACACTTCTGATA 4611

 792 .........AlaProGluProArgArg......GlnTyrThrLeuGluAs 803
             ||||||      ::::::      :::|||||:::   ||
4612 GAAAATTCAGCCCCCTATACTCATGAACATTTTCACTATACGGTGACTGA 4661

 803 pTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGlyT 820
     |   :::::::::   |||:::          :::      :::
4662 CATAAAAGATCTGACTAAACTAGGGGCCACTTATGACGATGCAAAG...A 4708

 820 hrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeu 836
       |||::::::      ||||||   :::::::|||   :::      :::
4709 AGTGTTGGGTTTATCAGGGAAAGCCTGTAATGCCTGATCAATTCACCTTT 4758

 837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLe 853
     |||   ::::::   :::|||:::|||||||||||||:::   :::
4759 GAACTATTAGATTTTCTTCATCAATTGACCCACCTCAGTTTCTCAAAAAC 4808

 853 uGlnGlnLeuValArgThrSer......ProTyrHisValLeuArgLeuP 868
      :::   |||::::::   :::      ||||||::::::|||:::
4809 AAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTACATGCTGAACCGGG 4858

 868 roGlyValAlaAspSerValValLysHisCysValProCysGlnLeuVal 884
       :::      ::::::::::::   |||      |||      |||
4859 ATCGAACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAGGTC 4908

 885 AsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySerHi 901
     |||||||:::   |||   :::      |||   |||:::||||||   ::
4909 AATGCCAGCAAGTCTGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCG 4958

 901 sProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLysT 918
     :||||||:::|||||||||:::|||||||||||||||||||||:::   |
4959 ACCCGGCACCCACTGGGAAATTGATTTCACTGAGGTAAAACCTGGCCTGT 5008

 918 yrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpVal 934
     |||||   ||||||||||||||||||:::||||||||||||||||||||
5009 ATGGGTATAAATATCTTTTAGTTTTCATAGACACTTTCTCTGGATGGGTA 5058

 935 GluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLysIl 951
     ||||||:::|||||||||||||||||:::   ||||||:::|||||||::
5059 GAAGCTTTCCCAACCAAGAAAGAAACTGCCAAAGTTGTAACCAAGAAGCT 5108

 951 eLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerA 968
     :|||||||||||||||||||||||||||:::|||:::|||:::|||:::|
5109 ACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAGGTATTGGGAACCG 5158

 968 spAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIle 984
     |||||||||||||||||||||:::::|||||||||   :::|||   :::
5159 ACAATGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGATTTA 5208

 985 LeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerGl 1001
     ||||||:::||||||||||||||||||||||||||||||||||||||||||
5209 CTGGGGGTTGATTGGAAACTACATTGTGCTTACAGACCCCAGAGTTCAGG 5258

1001 yGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuT 1018
     ||||||||||||||||||||||||||||||||||||||||||||||||||
5259 TCAGGTAGAAAGAATGAATAGGACAATCAAGGAGACTTTAACTAAATTGA 5308

1018 hrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeu 1034
```

```
          ||      ||||||   :::||||||:::   |||||||||:::::|||
       5309 CGCTTGCAACTGGCTCTAGGGACTGGGTGCTCCTGCTTCCCCTAGCCCTG 5358

       1035 PheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGluLe 1051
            :::|||:::|||||||||||||||   ||||||||||||||||||::
       5359 TATCGAGCCCGCAACACGCCGGGCCCCCATGGTCTCACCCCATATGAAAT 5408

       1051 uLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisSerA 1068
            :|||||||||:::|||||||||||||||:::::::::                |
       5409 CTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGATCCTGACATGG 5458

       1068 laAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGlu 1084
            ||   |||      :::      |||   ::::::|||:::||||||
       5459 CAAAGGTTACTCATAACCCCTCTCTCCAAGCCCATTTACAGGCACTCTAC 5508

       1085 TrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGlyGl 1101
               |||:::::::::::|||:::   |||      |||||:::
       5509 CTGGTCCAGCACGAAGTCTGGAGACCGTTGGCGGCAGCTTACCAA...GA 5555

       1101 yAspLeuGln.........ValProHisArgPheGlnValGlyAspServ 1115
             :::|||:::         ||||||||   |||:::||||||||||:::|
       5556 ACAACTGGACCGGCCGGTAGTGCCTCACCCTTTCCGAGTCGGTGACACAG 5605

       1115 alTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGly 1131
            ||:::|||||||||||||:::::   |||||||||   ||||||||||||
       5606 TGTGGGTCCGCAGACACCAAACTAAAAATCTAGAACCCCGCTGGAAAGGA 5655

       1132 ProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIl 1148
            ||||||   ||||||||||||||||||||||:::|||||||:::|||||
       5656 CCTTATACCGTCCTACTGACTACCCCCACCGCTCTCAAAGTGGACGGCAT 5705

       1148 eProThrTrpIleHisAlaSerHisValLysProAla.............  1160
            |   :::|||||||||||||:::|||||||||   |||
       5706 TGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGACACCAGGATTG 5755

       1161 ..ProProProAspSerGlyTrpArgAlaGluLysThrGluAsnProLeu 1176
              ||||||   :::|||   ||||||:::::::::::::::::|||||||||
       5756 AGCCACCATCGGAATCGACATGGCGTGTTCAACGCTCTCAAAATCCCCTA 5805

       1177 LysLeuArgLeuHisArg 1182
            |||:::||||||   |||
       5806 AAGATAAGATTGACCCGC 5823

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-9

seq_documentation_block:
; Sequence 9, Application US/08110300A
; Patent No. 5643756
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,300A
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-110-300A-9


alignment_scores:
            Quality: 3980.00              Length:   1206
              Ratio:  3.992                Gaps:     12
 Percent Similarity:  82.670   Percent Identity: 62.687

alignment_block:
 US-09-171-553B-5 x US-08-110-300A-9/rev   ..

 Align seg 1/1  to reverse of: US-08-110-300A-9  from: 1  to: 10367

       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         |||   :::|||:::::|||   |||||||||||:::|||||||:::||
    5360 GGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAGAGT 5311

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerV 34
         |   ||||||||||||   |||||||||||||||||||:::|||||||
    5310 CGGGGGGCAACCCGTCACCTTCCTAGTGGATACTGGGGCCCAACACTCCG 5261

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
         |||||  |||      |||  |||:::::::|||:::::|||||||:::
    5260 TGCTGACCCAAAATCCTGGACCCCTAAGTGACAAGTCTGCCTGGGTCCAA 5211

      51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
         ||||||||||||   ::::::|||   |||||||||   |||   |||
    5210 GGGGCTACTGGAGGGAAGCGGTATCGCTGGACCACGGATCGCCGAGTGCA 5161

      67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
          |||:::::::|||:::||||||||||||||||||   :::|||:::||||
    5160 CCTAGCCACCGGTAAGGTCACCCATTCTTTCCTCCATGTACCAGATTGCC 5111

      84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
         ||   ||||||||||||||||||||||||||||:::   |||||||||
    5110 CCTATCCTCTGCTAGGAAGAGATTTGCTGACTAAACTAAAAGCCCAAATT 5061

     101 SerPheGlu...GlnGlyArgProGluValSerValAsnAsnLysProIl 116
         ||||||   :::|||         |||:::      :::::::|||::
    5060 CACTTTGAGGGATCAGGAGCTCAGGTTGTGGGACCAATGGGACAGCCCCT 5011

     116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
         :   ||||||||||||:::::::::|||||||||||||||::::::   :
    5010 GCAAGTGCTGACCCTAAACATAGAAGATGAGTATCGGCTACATGAGACCT 4961

     133 lnValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
         ::      ||||||      :::   |||   ||||||:::::::|||||
    4960 CAAAAGGGCCAGATGTGCCTCTAGGGTCCACATGGCTCTCTGATTTTCCC 4911

     149 GlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPr 165
         ||||||||||||||||:::|||||||||||||||   :::      ||
    4910 CAGGCCTGGGCAGAAACCGGGGGCATGGGGCTGGCCGTTCGCCAAGCTCC 4861

     165 oGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnT 182
         |   :::|||   |||||||||:::::|||||||||||||:::::||||
    4860 TCTGATCATACCTCTGAAGGCAACCTCTACCCCCGTGTCCATAAAACAAT 4811

     182 yrProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArg 198
         |||||:::|||:::||||||||||   ||||||   |||||||:::||||||
    4810 ACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGA 4761

     199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPr 215
         |||:::::|||||||||||||||||||   ||||||||||||||||||||
    4760 CTGCTGGATCAGGGAATTCTGGTACCCTGCCAGTCCCCCTGGAACACGCC 4711

     215 oLeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnA 232
         ||||||||||||:::|||||||||||||||||||||||||||||||||||
    4710 CCTGCTACCCGTTAAGAAACCGGGGACTAATGATTATAGGCCTGTCCAGG 4661

     232 spLeuArgGluValAsnLysArgValGlnAspIleHisProThrValPro 248
         ||||||||||||||||||||||||||:::||||||||||||||||||||||
    4660 ATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCC 4611

     249 AsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrTh 265
         ||||||||||||||||||||||:::||||||||||:::::::::|||||||
    4610 AACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACAC 4561

     265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4560 TGTGCTTGACTTAAAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCA 4511

     282 erGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
         |||||   ||||||||||||||||||||||||||||||      |||   :::
    4510 GTCAGTCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA 4461

     299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4460 GGACAATTAACCTGGACCAGACTCCCGCAGGGTTTCAAAAACAGTCCCAC 4411

     315 rIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
         |:::|||||||||||||||||||||||||||||:::|||||||||||||
    4410 CCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGGATCCAGC 4361

     332 isProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGly 348
         |||||::::::   ||||||||||||||||||||||||||||||||:::
    4360 ACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCC 4311

     349 AlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLe 365
         :::::::::   ||||||   :::||||||::::|||||||||      ||
    4310 ACTTCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCT 4261

     365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
         |:::|||||||||||||||||||||||||||||||||||||||||||::::
    4260 AGGGGACCTCGGATATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGA 4211

     382 rgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
         :::::|||   ||||||||||||   ||||::   ||||||||||||||||
    4210 AACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTG 4161

     399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
         ||||||||||||||:::|||||||:::      |||:::|||   |||
    4160 ACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 4111

     415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
          :::|||:::||||||||||||||||||||||||||||||||||||||||
    4110 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGA 4061

     432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
         ||||||||||||   :::||||||||||||||||||||||||||||||||
    4060 TCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACG 4011

     449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIl 465
                 |||:::|||:::|||::::::|||||||||::::::   ||
    4010 GGGACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGAT 3961

     465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
         ||||:::|||||||||:::|||||||||||:::|||||||||:::||||
    3960 CAAGCAGGCTCTCTTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTA 3911

     482 ysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
         ||||||||   |||:::|||||||||:::::|||   |||:::||||||
    3910 AGCCCTTCGAACTTTTTGTTGACGAGAAGCAGGGCTACGCCAAAGGTGTC 3861

     499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLy 515
```

```
    1101 yAspLeuGln.........ValProHisArgPheGlnValGlyAspSerV 1115
         :::|||:::         |||||||||   |||:::||||||||||:::|
    2040 GCAACTAGACCGGCCGGTAGTACCTCACCCTTTCCGGGTCGGTGACACAG 1991

    1115 alTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGly 1131
         ||:::|||||||||||||:::::   |||||||||   ||||||||||||
    1990 TGTGGGTCCGCAGACACCAAACTAAAAATCTAGAACCCCGCTGGAAAGGA 1941

    1132 ProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIl 1148
         ||||||   ||||||||||||||||||||||:::|||||||:::|||||
    1940 CCCTATACCGTCCTACTGACTACCCCCACCGCTCTCAAAGTAGACGGCAT 1891

    1148 eProThrTrpIleHisAlaSerHisValLysProAla..............  1160
         |   :::||||||||||||:::|||||||||   |||
    1890 TGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGACACCAGGATTG 1841

    1161 ..ProProProAspSerGlyTrpArgAlaGluLysThrGluAsnProLeu 1176
           ||||||   :::|||   ||||||:::::::::::::::|||||||||
    1840 AGCCACCAGCAGAATCGACATGGCGTGTTCAACGCTCCCAAAATCCCCTA 1791

    1177 LysLeuArgLeuHisArg 1182
         |||:::||||||   |||
    1790 AAGATAAGATTGACCCGC 1773

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-9

seq_documentation_block:
; Sequence 9, Application US/08886642
; Patent No. 5952474
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  45 Rockefeller Plaza, Suite 2800
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/886,642
;      FILING DATE:  01-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/110,300
;      FILING DATE:  20-AUG-1993
;      APPLICATION NUMBER:  07/938,100
;      FILING DATE:  28-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hone, William J.
;      REGISTRATION NUMBER:  26,739
;      REFERENCE/DOCKET NUMBER:  07763/010002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212/765-5070
;      TELEFAX:  212/258-2291
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-886-642-9


alignment_scores:

            Quality: 3980.00              Length:   1206
              Ratio:  3.992                 Gaps:     12
Percent Similarity:  82.670   Percent Identity: 62.687

alignment_block:
 US-09-171-553B-5 x US-08-886-642-9/rev   ..

 Align seg 1/1  to reverse of: US-08-886-642-9  from: 1  to: 10367

       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         |||   :::|||:::::::|||   |||||||||||:::|||||:::||
    5360 GGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAGAGT 5311

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerV 34
         |   |||||||||||   |||||||||||||||||||||:::|||||||
    5310 CGGGGGGCAACCCGTCACCTTCCTAGTGGATACTGGGGCCCAACACTCCG 5261

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
         |||||   |||      |||   |||::::::|||::::::|||||:::
    5260 TGCTGACCCAAAATCCTGGACCCCTAAGTGACAAGTCTGCCTGGGTCCAA 5211

      51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
         ||||||||||||   ::::::|||   |||||||||   |||   |||
    5210 GGGGCTACTGGAGGGAAGCGGTATCGCTGGACCACGGATCGCCGAGTGCA 5161

      67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
         |||:::::::|||:::|||||||||||||||||||   :::|||:::||||
    5160 CCTAGCCACCGGTAAGGTCACCCATTCTTTCCTCCATGTACCAGATTGCC 5111

      84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
         ||   |||||||||||||||||||||||||||||:::   |||||||||
    5110 CCTATCCTCTGCTAGGAAGAGATTTGCTGACTAAACTAAAAGCCCAAATT 5061

     101 SerPheGlu...GlnGlyArgProGluValSerValAsnAsnLysProIl 116
         ||||||   :::|||         |||:::      :::::|||::
    5060 CACTTTGAGGGATCAGGAGCTCAGGTTGTGGGACCAATGGGACAGCCCCT 5011

     116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
         :   ||||||||||||:::::::::|||||||||||||||||:::::   :
    5010 GCAAGTGCTGACCCTAAACATAGAAGATGAGTATCGGCTACATGAGACCT 4961

     133 lnValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
         ::     ||||||      :::   |||   ||||||:::::|||||
    4960 CAAAAGGGCCAGATGTGCCTCTAGGGTCCACATGGCTCTCTGATTTTCCC 4911

     149 GlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPr 165
         ||||||||||||||||:::|||||||||||||||   :::      ||
    4910 CAGGCCTGGGCAGAAACCGGGGGCATGGGGCTGGCCGTTCGCCAAGCTCC 4861

     165 oGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnT 182
         |   :::|||   |||||||||:::::::||||||||||||:::::||||
    4860 TCTGATCATACCTCTGAAGGCAACCTCTACCCCCGTGTCCATAAAACAAT 4811

     182 yrProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArg 198
         |||||:::|||:::|||||||||   ||||||   ||||||:::||||||
    4810 ACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGA 4761

     199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPr 215
         |||:::::|||||||||||||||||   ||||||||||||||||||||
    4760 CTGCTGGATCAGGGAATTCTGGTACCCTGCCAGTCCCCCTGGAACACGCC 4711

     215 oLeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnA 232
         ||||||||||||:::|||||||||||||||||||||||||||||||||||
    4710 CCTGCTACCCGTTAAGAAACCGGGGACTAATGATTATAGGCCTGTCCAGG 4661

     232 spLeuArgGluValAsnLysArgValGlnAspIleHisProThrValPro 248
         ||||||||||||||||||||||||:::|||||||||||||||||||||||
    4660 ATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCC 4611

     249 AsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrTh 265
         ||||||||||||||||||:::|||||||||:::::::::|||||||||
    4610 AACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACAC 4561
```

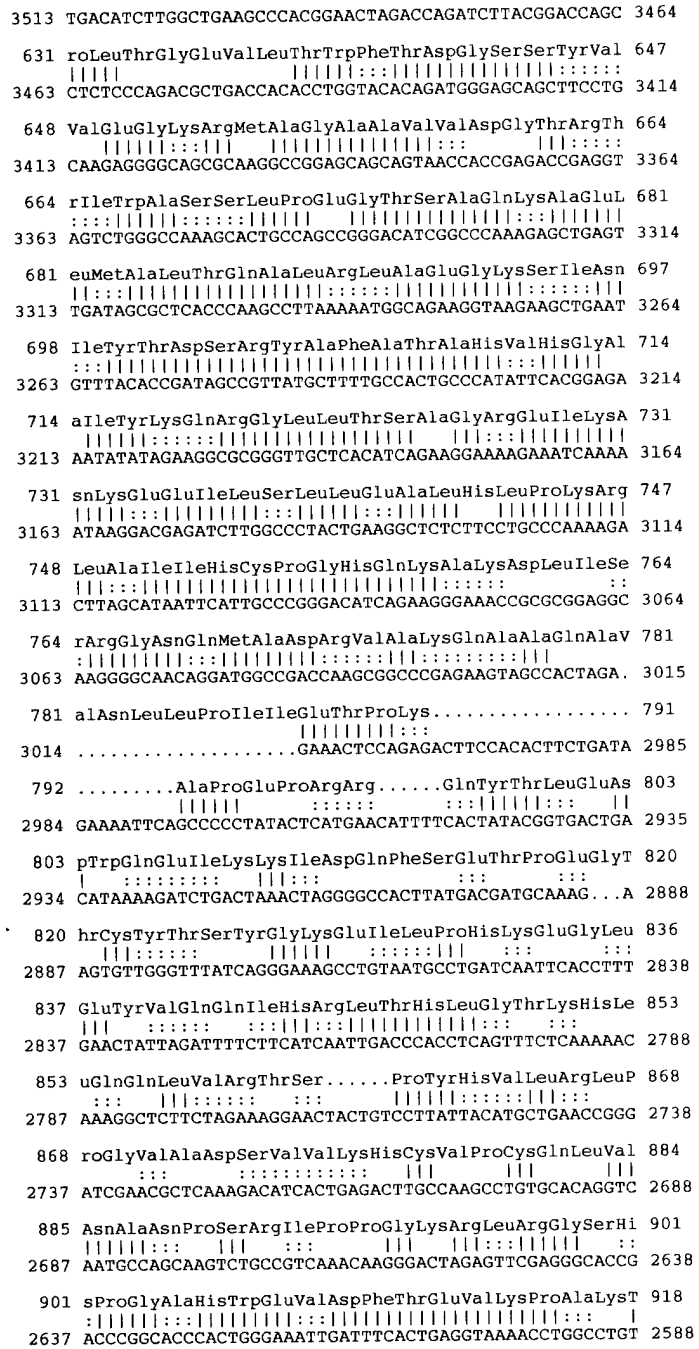

265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
4560 TGTGCTTGACTTAAAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCA 4511
282 erGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
4510 GTCAGTCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA 4461
299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
4460 GGACAATTAACCTGGACCAGACTCCCGCAGGGTTTCAAAAACAGTCCCAC 4411
315 rIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
4410 CCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGGATCCAGC 4361
332 isProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGly 348
4360 ACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCC 4311
349 AlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLe 365
4310 ACTTCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCT 4261
365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
4260 AGGGGACCTCGGATATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGA 4211
382 rgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
4210 AACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTG 4161
399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
4160 ACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 4111
415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
4110 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGA 4061
432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
4060 TCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACG 4011
449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIl 465
4010 GGGACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGAT 3961
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
3960 CAAGCAGGCTCTCTTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTA 3911
482 ysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
3910 AGCCCTTCGAACTTTTGTTGACGAGAAGCAGGGCTACGCCAAAGGTGTC 3861
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLy 515
3860 CTAACGCAAAAACTGGGGCCTTGGCGTCGGCCGGTGGCCTACCTGTCCAA 3811
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
3810 AAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAG 3761
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
3760 CAGCCATCGCCGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAG 3711
549 AsnIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPr 565
3710 CCACTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTTAAGCAACC 3661
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
3660 CCCTGATCGCTGGCTCTCCAACGCCCGAATGACCCACTACCAGGCTCTGC 3611
582 euLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
3610 TTCTGGACACGGACCGAGTCCAGTTCGGACCAATAGTGGCCCTAAACCCA 3561
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHi 614
3560 GCTACGCTGCTCCCT...CTACCTGAGGAGGGGCTGCAACATGACTGCCT 3514
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
3513 TGACATCTTGGCTGAAGCCCACGGAACTAGACCAGATCTTACGGACCAGC 3464
631 roLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrVal 647
3463 CTCTCCCAGACGCTGACCACACCTGGTACACAGATGGGAGCAGCTTCCTG 3414
648 ValGluGlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgTh 664
3413 CAAGAGGGGCAGCGCAAGGCCGGAGCAGCAGTAACCACCGAGACCGAGGT 3364
664 rIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluL 681
3363 AGTCTGGGCCAAAGCACTGCCAGCCGGGACATCGGCCCAAAGAGCTGAGT 3314
681 euMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
3313 TGATAGCGCTCACCCAAGCCTTAAAAATGGCAGAAGGTAAGAAGCTGAAT 3264
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
3263 GTTTACACCGATAGCCGTTATGCTTTTGCCACTGCCCATATTCACGGAGA 3214
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
3213 AATATATAGAAGGCGCGGGTTGCTCACATCAGAAGGAAAAGAAATCAAAA 3164
731 snLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
3163 ATAAGGACGAGATCTTGGCCCTACTGAAGGCTCTCTTCCTGCCCAAAAGA 3114
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSe 764
3113 CTTAGCATAATTCATTGCCCGGGACATCAGAAGGGAAACCGCGCGGAGGC 3064
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaV 781
3063 AAGGGGCAACAGGATGGCCGACCAAGCGGCCCGAGAAGTAGCCACTAGA. 3015
781 alAsnLeuLeuProIleIleGluThrProLys..... 791
3014 .....GAAACTCCAGAGACTTCCACACTTCTGATA 2985
792 .....AlaProGluProArgArg......GlnTyrThrLeuGluAs 803
2984 GAAAATTCAGCCCCCTATACTCATGAACATTTTCACTATACGGTGACTGA 2935
803 pTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGlyT 820
2934 CATAAAAGATCTGACTAAACTAGGGGCCACTTATGACGATGCAAAG...A 2888
820 hrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeu 836
2887 AGTGTTGGGTTTATCAGGGAAAGCCTGTAATGCCTGATCAATTCACCTTT 2838
837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLe 853
2837 GAACTATTAGATTTTCTTCATCAATTGACCCACCTCAGTTTCTCAAAAAC 2788
853 uGlnGlnLeuValArgThrSer......ProTyrHisValLeuArgLeuP 868

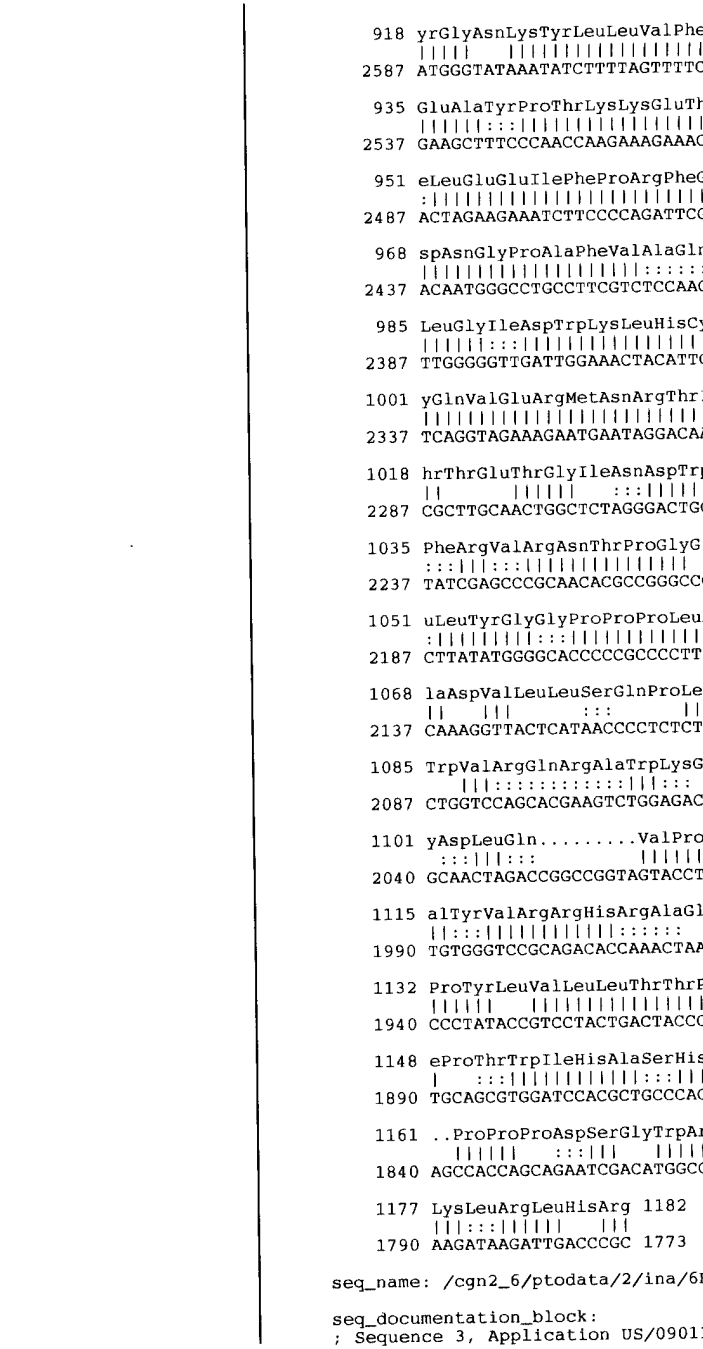
```
2787 AAAGGCTCTTCTAGAAAGGAACTACTGTCCTTATTACATGCTGAACCGGG 2738

 868 roGlyValAlaAspSerValValLysHisCysValProCysGlnLeuVal 884
        :::      ::::::::::::   |||      |||      |||
2737 ATCGAACGCTCAAAGACATCACTGAGACTTGCCAAGCCTGTGCACAGGTC 2688

 885 AsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySerHi 901
     ||||||:::   |||   :::      |||   |||:::||||||   ::
2687 AATGCCAGCAAGTCTGCCGTCAAACAAGGGACTAGAGTTCGAGGGCACCG 2638

 901 sProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLysT 918
     :|||||||:::|||||||||:::|||||||||||||||||||||||:::   |
2637 ACCCGGCACCCACTGGGAAATTGATTTCACTGAGGTAAAACCTGGCCTGT 2588

 918 yrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpVal 934
     |||||   ||||||||||||||||||:::|||||||||||||||||||||
2587 ATGGGTATAAATATCTTTTAGTTTTCATAGACACTTTCTCTGGATGGGTA 2538

 935 GluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLysIl 951
     ||||||:::|||||||||||||||||||:::   ||||||:::||||||::
2537 GAAGCTTTCCCAACCAAGAAAGAAACTGCCAAAGTTGTAACCAAGAAGCT 2488

 951 eLeuGluGluIleProProArgPheGlyIleProLysValIleGlySerA 968
     :|||||||||||||||||||||||||||:::|||:::|||:::|||:::|
2487 ACTAGAAGAAATCTTCCCCAGATTCGGCATGCCACAGGTATTGGGAACCG 2438

 968 spAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIle 984
     ||||||||||||||||||||:::::|||||||||   :::|||   :::
2437 ACAATGGGCCTGCCTTCGTCTCCAAGGTAAGTCAGACAGTAGCCGATTTA 2388

 985 LeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerGl 1001
     ||||||:::||||||||||||||||||||||||||||||||||||||||||
2387 TTGGGGGTTGATTGGAAACTACATTGTGCTTACAGACCCCAGAGTTCAGG 2338

1001 yGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuT 1018
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2337 TCAGGTAGAAAGAATGAATAGGACAATCAAGGAGACTTTAACTAAATTGA 2288

1018 hrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeu 1034
     ||      ||||||   :::|||||||:::   |||||||||:::::|||
2287 CGCTTGCAACTGGCTCTAGGGACTGGGTGCTCCTGCTTCCCCTTGCCCTG 2238

1035 PheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGluLe 1051
     :::|||:::|||||||||||||||||      ||||||||||||||||||::
2237 TATCGAGCCCGCAACACGCCGGGCCCCCACGGTCTCACCCCATATGAAAT 2188

1051 uLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisSerA 1068
     :|||||||||:::|||||||||||||:::::::::                |
2187 CTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGATCCTGACATGG 2138

1068 laAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGlu 1084
     ||   |||      :::      |||   ::::::|||:::||||||
2137 CAAAGGTTACTCATAACCCCTCTCTCCAAGCTCATTTACAGGCACTCTAC 2088

1085 TrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGlyGl 1101
        |||:::::::::::::|||:::   |||      |||||||:::
2087 CTGGTCCAGCACGAAGTCTGGAGACCGTTGGCGGCAGCTTACCAA...GA 2041

1101 yAspLeuGln.........ValProHisArgPheGlnValGlyAspSerV 1115
      :::|||:::         |||||||||   |||:::|||||||||:::|
2040 GCAACTAGACCGGCCGGTAGTACCTCACCCTTTCCGGGTCGGTGACACAG 1991

1115 alTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGly 1131
     ||:::|||||||||||||:::::   |||||||||   ||||||||||||
1990 TGTGGGTCCGCAGACACCAAACTAAAAATCTAGAACCCCGCTGGAAAGGA 1941

1132 ProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIl 1148
     ||||||   ||||||||||||||||||||||||:::|||||||:::|||||
1940 CCCTATACCGTCCTACTGACTACCCCCACCGCTCTCAAAGTAGACGGCAT 1891

1148 eProThrTrpIleHisAlaSerHisValLysProAla..... 1160
     |   :::|||||||||||||:::|||||||||   |||
1890 TGCAGCGTGGATCCACGCTGCCCACGTAAAGGCTGCCGACACCAGGATTG 1841
```

```
1161 ..ProProProAspSerGlyTrpArgAlaGluLysThrGluAsnProLeu 1176
       ||||||   :::|||   ||||||:::::::::::::::|||||||||
1840 AGCCACCAGCAGAATCGACATGGCGTGTTCAACGCTCCCAAAATCCCCTA 1791

1177 LysLeuArgLeuHisArg 1182
     |||:::||||||   |||
1790 AAGATAAGATTGACCCGC 1773
```

```
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-9

seq_documentation_block:
; Sequence 9, Application PC/TUS9308041
;  GENERAL INFORMATION:
;    APPLICANT:  The Public Health Research Institute of the City of
;    APPLICANT:  New York, Inc.
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08041
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/938,100
;      FILING DATE:  28-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/_____
;      FILING DATE:  20-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-08041-9


alignment_scores:
            Quality: 3980.00              Length:   1206
              Ratio:  3.992                 Gaps:     12
 Percent Similarity:  82.670   Percent Identity: 62.687

alignment_block:
 US-09-171-553B-5 x PCT-US93-08041-9/rev   ..

 Align seg 1/1  to reverse of: PCT-US93-08041-9  from: 1  to: 10367

       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         |||   :::|||:::::::|||   |||||||||||:::|||||||:::||
    5360 GGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAGAGT 5311
```

17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerV 34
5310 CGGGGGGCAACCCGTCACCTTCCTAGTGGATACTGGGGCCCAACACTCCG 5261
34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
5260 TGCTGACCCAAAATCCTGGACCCCTAAGTGACAAGTCTGCCTGGGTCCAA 5211
51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
5210 GGGGCTACTGGAGGGAAGCGGTATCGCTGGACCACGGATCGCCGAGTGCA 5161
67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
5160 CCTAGCCACCGGTAAGGTCACCCATTCTTTCCTCCATGTACCAGATTGCC 5111
84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
5110 CCTATCCTCTGCTAGGAAGAGATTTGCTGACTAAACTAAAAGCCCAAATT 5061
101 SerPheGlu...GlnGlyArgProGluValSerValAsnAsnLysProIl 116
5060 CACTTTGAGGGATCAGGAGCTCAGGTTGTGGGACCAATGGGACAGCCCCT 5011
116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
5010 GCAAGTGCTGACCCTAAACATAGAAGATGAGTATCGGCTACATGAGACCT 4961
133 lnValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
4960 CAAAAGGGCCAGATGTGCCTCTAGGGTCCACATGGCTCTCTGATTTTCCC 4911
149 GlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPr 165
4910 CAGGCCTGGGCAGAAACCGGGGGCATGGGGCTGGCCGTTCGCCAAGCTCC 4861
165 oGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnT 182
4860 TCTGATCATACCTCTGAAGGCAACCTCTACCCCCGTGTCCATAAAACAAT 4811
182 yrProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArg 198
4810 ACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGA 4761
199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPr 215
4760 CTGCTGGATCAGGGAATTCTGGTACCCTGCCAGTCCCCCTGGAACACGCC 4711
215 oLeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnA 232
4710 CCTGCTACCCGTTAAGAAACCGGGGACTAATGATTATAGGCCTGTCCAGG 4661
232 spLeuArgGluValAsnLysArgValGlnAspIleHisProThrValPro 248
4660 ATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCC 4611
249 AsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrTh 265
4610 AACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACAC 4561
265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
4560 TGTGCTTGACTTAAAAGATGCTTTTTTCTGCCTGAGACTCCACCCCACCA 4511
282 erGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
4510 GTCAGTCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA 4461
299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
4460 GGACAATTAACCTGGACCAGACTCCCGCAGGGTTTCAAAAACAGTCCCAC 4411
315 rIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
4410 CCTGTTTGATGAAGCCCTGCACAGGGACCTCGCAGACTTCCGGATCCAGC 4361
332 isProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGly 348
4360 ACCCAGACCTGATTCTGCTCCAGTATGTAGATGACTTACTGCTGGCCGCC 4311
349 AlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLe 365
4310 ACTTCTGAGCTTGACTGTCAACAAGGTACGCGGGCCCTGTTACAAACCCT 4261
365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
4260 AGGGGACCTCGGATATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGA 4211
382 rgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
4210 AACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTG 4161
399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
4160 ACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 4111
415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
4110 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGA 4061
432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
4060 TCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACG 4011
449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIl 465
4010 GGGACTCTGTTTGAGTGGGGCCCAGACCAGCAAAAGGCCTACCAAGAGAT 3961
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
3960 CAAGCAGGCTCTCTTAACTGCCCCTGCCCTGGGATTGCCAGACTTGACTA 3911
482 ysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
3910 AGCCCTTCGAACTTTTTGTTGACGAGAAGCAGGGCTACGCCAAAGGTGTC 3861
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLy 515
3860 CTAACGCAAAAACTGGGGCCTTGGCGTCGGCCGGTGGCCTACCTGTCCAA 3811
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
3810 AAAGCTAGACCCAGTGGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAG 3761
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
3760 CAGCCATCGCCGTTCTGACCAAAGACGCTGGCAAGCTCACCATGGGACAG 3711
549 AsnIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPr 565
3710 CCACTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTTAAGCAACC 3661
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
3660 CCCTGATCGCTGGCTCTCCAACGCCCGAATGACCCACTACCAGGCTCTGC 3611
582 euLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
3610 TTCTGGACACGGACCGAGTCCAGTTCGGACCAATAGTGGCCCTAAACCCA 3561
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHi 614
3560 GCTACGCTGCTCCCT...CTACCTGAGGAGGGGCTGCAACATGACTGCCT 3514
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631



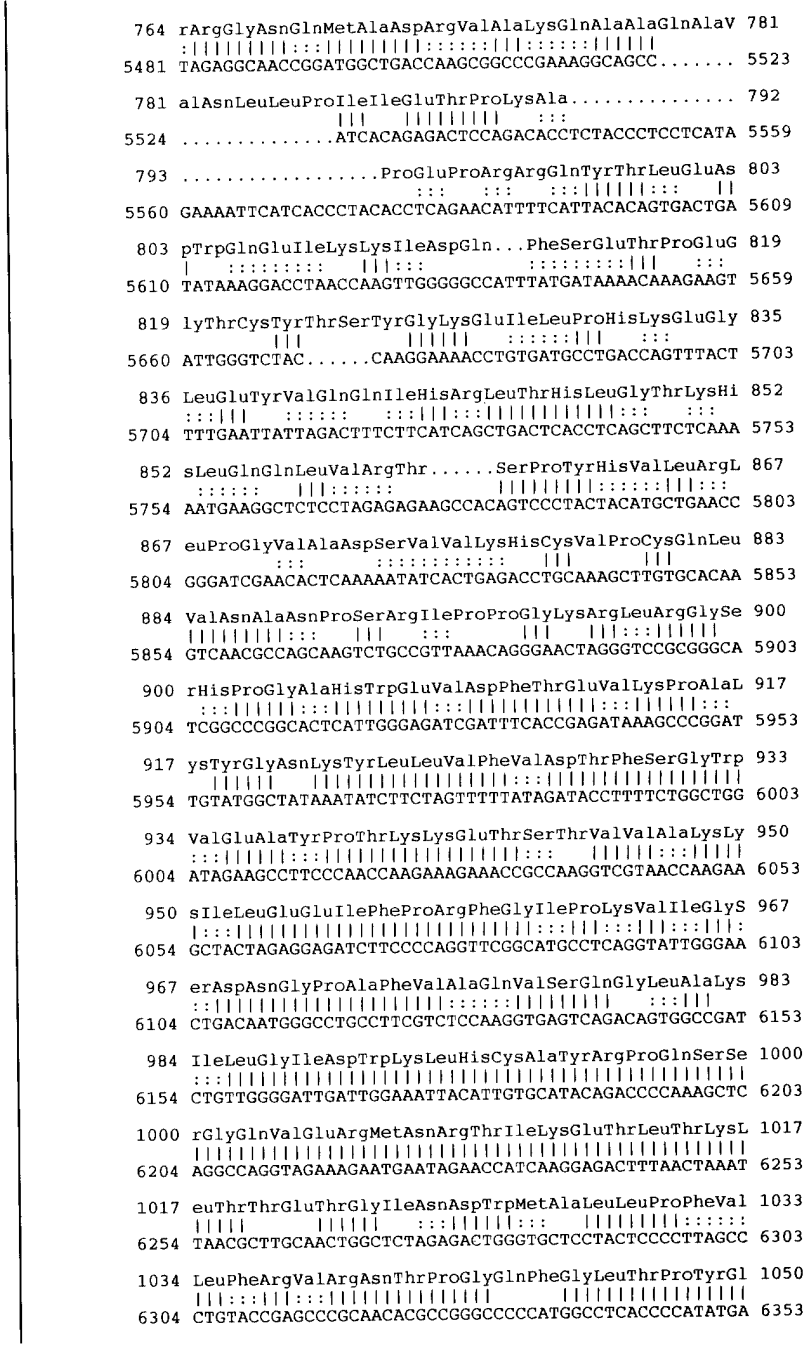


seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-3

seq_documentation_block:
; Sequence 3, Application US/09011745

```
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 7308
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
US-09-011-745-3


alignment_scores:
            Quality: 3967.00              Length:   1210
              Ratio:  3.959                Gaps:     13
 Percent Similarity:  82.810   Percent Identity: 62.479

alignment_block:
 US-09-171-553B-5 x US-09-011-745-3   ..

 Align seg 1/1  to: US-09-011-745-3  from: 1  to: 7308

       1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
         |||   :::|||::::::|||   |||||||||||||:::|||||||||||
    3184 GGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGT 3233

      17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisServ 34
         |   ||||||||||||   |||||||||||||||||||:::|||||||
    3234 CGGGGGGCAACCCGTCACCTTCCTGGTAGATACTGGGGCCCAACACTCCG 3283

      34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
         |||||   |||       |||   |||:::::|||:::::||||||:::
    3284 TGCTGACCCAAAATCCTGGACCCCTAAGTGATAAGTCTGCCTGGGTCCAA 3333

      51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
         ||||||||||||   ::::::|||   |||||||||   |||   |||
    3334 GGGGCTACTGGAGGAAAGCGGTATCGCTGGACCACGGATCGCAAAGTACA 3383

      67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
          |||:::::::|||:::|||||||||||||||||||   :::|||:::||||
    3384 TCTAGCTACCGGTAAGGTCACCCACTCTTTCCTCCATGTACCAGACTGTC 3433

      84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
         ||   |||||||||||||||||||||||||||||:::   |||||||||
    3434 CCTATCCTCTGTTAGGAAGAGATTTGCTGACTAAACTAAAAGCCCAAATC 3483

     101 SerPheGlu...GlnGlyArgProGluValSerValAsnAsnLysProIl 116
         ||||||   :::|||         ::::::      ::::::|||::
    3484 CACTTTGAGGGATCAGGAGCTCAGGTTATGGGACCAATGGGGCAGCCCCT 3533

     116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
         :   ||||||||||||:::::::::|||||||:::|||||||::::::   :
    3534 GCAAGTGTTGACCCTAAATATAGAAGATGAGCATCGGCTACATGAGACCT 3583

     133 lnValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
         ::   :::|||||||   ::::::   |||   |||||||:::::|||||
    3584 CAAAAGAGCCAGATGTTTCTCTAGGGTCCACATGGCTGTCTGATTTTCCT 3633

     149 GlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPr 165

         ||||||||||||||||||:::|||||||||||||||   :::      ||
    3634 CAGGCCTGGGCGGAAACCGGGGGCATGGGACTGGCAGTTCGCCAAGCTCC 3683

     165 oGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnT 182
         |   :::|||   ||||||||||:::::|||||||||||||:::::||||
    3684 TCTGATCATACCTCTGAAAGCAACCTCTACCCCCGTGTCCATAAAACAAT 3733

     182 yrProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArg 198
         |||||:::|||:::|||||||||   ||||||   ||||||:::|||||||
    3734 ACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGA 3783

     199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPr 215
         |||::::::|||||||||||||||||||   |||||||||||||||||||
    3784 CTGTTGGACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCC 3833

     215 oLeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnA 232
         ||||||||||||:::|||||||||||||||||||||||||||||||||||
    3834 CCTGCTACCCGTTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGG 3883

     232 spLeuArgGluValAsnLysArgValGlnAspIleHisProThrValPro 248
         ||||||||||||||||||||||||:::|||||||||||||||||||||||
    3884 ATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCC 3933

     249 AsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrTh 265
         ||||||||||||||||||||:::|||||||||:::::::::|||||||||
    3934 AACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACAC 3983

     265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3984 TGTGCTTGATTTAAAGGATGCCTTTTTCTGCCTGAGACTCCACCCCACCA 4033

     282 erGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
         ||||||||||||||||||||||||||||||||   |||   :::
    4034 GTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA 4083

     299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4084 GGACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCAC 4133

     315 rIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
         |:::|||||||||||||||||||||||||||||:::|||||||||||||||
    4134 CCTGTTTGATGAGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGC 4183

     332 isProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGly 348
         |||||::::::   |||||||||||||||||||||||||||||||||:::
    4184 ACCCAGACTTGATCCTGCTACAGTACGTGGATGACTTACTGCTGGCCGCC 4233

     349 AlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLe 365
         :::::::::   ||||||   :::|||||||::|||||||||      ||
    4234 ACTTCTGAGCTAGACTGCCAACAAGGTACTCGGGCCCTGTTACAAACCCT 4283

     365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
         |::::::|||||||||||||||||||||||||||||||||||||||||::::
    4284 AGGGAACCTCGGGTATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGA 4333

     382 rgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
         :::::|||   ||||||||||||   |||:::   ||||||||||||||||
    4334 AACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTG 4383

     399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
         ||||||||||||||||:::|||||||:::      |||:::|||   |||
    4384 ACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 4433

     415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
         :::|||:::|||||||||||||||||||||||||||||||||||||||||
    4434 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGA 4483

     432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
         ||||||||||||||   :::|||||||||||||||||||||||||||||
    4484 TCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACG 4533

     449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIl 465
                 |||:::|||:::|||::::::|||||||||::::::   ||
```

4534 GGGACTCTGTTTAATTGGGGCCCAGACCAACAAAAGGCCTATCAAGAAAT 4583
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
4584 CAAGCAAGCTCTTCTAACTGCCCCAGCCCTGGGGTTGCCAGATTTGACTA 4633
482 ysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
4634 AGCCCTTTGAACTCTTTGTCGACGAGAAGCAGGGCTACGCCAAAGGTGTC 4683
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLy 515
4684 CTAACGCAAAAACTGGGACCTTGGCGTCGGCCGGTGGCCTACCTGTCCAA 4733
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
4734 AAAGCTAGACCCAGTAGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAG 4783
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
4784 CAGCCATTGCCGTACTGACAAAGGATGCAGGCAAGCTAACCATGGGACAG 4833
549 AsnIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPr 565
4834 CCACTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTCAAACAACC 4883
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
4884 CCCCGACCGCTGGCTTTCCAACGCCCGGATGACTCACTATCAGGCCTTGC 4933
582 euLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
4934 TTTTGGACACGGACCGGGTCCAGTTCGGACCGGTGGTAGCCCTGAACCCG 4983
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHi 614
4984 GCTACGCTGCTCCCA...CTGCCTGAGGAAGGGCTGCAACACAACTGCCT 5030
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
5031 TGATATCCTGGCCGAAGCCCACGGAACCCGACCCGACCTAACGGACCAGC 5080
631 roLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrVal 647
5081 CGCTCCCAGACGCCGACCACACCTGGTACACGGATGGAAGCAGTCTCTTA 5130
648 ValGluGlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgTh 664
5131 CAAGAGGGACAGCGTAAGGCGGGAGCTGCGGTGACCACCGAGACCGAGGT 5180
664 rIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluL 681
5181 AATCTGGGCTAAAGCCCTGCCAGCCGGGACATCCGCTCAGCGGGCTGAAC 5230
681 euMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
5231 TGATAGCACTCACCCAGGCCCTAAAGATGGCAGAAGGTAAGAAGCTAAAT 5280
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
5281 GTTTATACTGATAGCCGTTATGCTTTTGCTACTGCCCATATCCATGGAGA 5330
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
5331 AATATACAGAAGGCGTGGGTTGCTCACATCAGAAGGCAAAGAGATCAAAA 5380
731 snLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
5381 ATAAAGACGAGATCTTGGCCCTACTAAAAGCCCTCTTTCTGCCCAAAAGA 5430
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSe 764
5431 CTTAGCATAATCCATTGTCCAGGACATCAAAAGGGACACAGCGCCGAGGC 5480


764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaV 781
5481 TAGAGGCAACCGGATGGCTGACCAAGCGGCCCGAAAGGCAGCC..... 5523
781 alAsnLeuLeuProIleIleGluThrProLysAla..... 792
5524 .....ATCACAGAGACTCCAGACACCTCTACCCTCCTCATA 5559
793 .....ProGluProArgArgGlnTyrThrLeuGluAs 803
5560 GAAAATTCATCACCCTACACCTCAGAACATTTTCATTACACAGTGACTGA 5609
803 pTrpGlnGluIleLysLysIleAspGln...PheSerGluThrProGluG 819
5610 TATAAAGGACCTAACCAAGTTGGGGGCCATTTATGATAAAACAAAGAAGT 5659
819 lyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGly 835
5660 ATTGGGTCTAC......CAAGGAAAACCTGTGATGCCTGACCAGTTTACT 5703
836 LeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHi 852
5704 TTTGAATTATTAGACTTTCTTCATCAGCTGACTCACCTCAGCTTCTCAAA 5753
852 sLeuGlnGlnLeuValArgThr......SerProTyrHisValLeuArgL 867
5754 AATGAAGGCTCTCCTAGAGAGAAGCCACAGTCCCTACTACATGCTGAACC 5803
867 euProGlyValAlaAspSerValValLysHisCysValProCysGlnLeu 883
5804 GGGATCGAACACTCAAAAATATCACTGAGACCTGCAAAGCTTGTGCACAA 5853
884 ValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySe 900
5854 GTCAACGCCAGCAAGTCTGCCGTTAAACAGGGAACTAGGGTCCGCGGGCA 5903
900 rHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaL 917
5904 TCGGCCCGGCACTCATTGGGAGATCGATTTCACCGAGATAAAGCCCGGAT 5953
917 ysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrp 933
5954 TGTATGGCTATAAATATCTTCTAGTTTTTATAGATACCTTTTCTGGCTGG 6003
934 ValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLy 950
6004 ATAGAAGCCTTCCCAACCAAGAAAGAAACCGCCAAGGTCGTAACCAAGAA 6053
950 sIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlyS 967
6054 GCTACTAGAGGAGATCTTCCCCAGGTTCGGCATGCCTCAGGTATTGGGAA 6103
967 erAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLys 983
6104 CTGACAATGGGCCTGCCTTCGTCTCCAAGGTGAGTCAGACAGTGGCCGAT 6153
984 IleLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSe 1000
6154 CTGTTGGGGATTGATTGGAAATTACATTGTGCATACAGACCCCAAAGCTC 6203
1000 rGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysL 1017
6204 AGGCCAGGTAGAAAGAATGAATAGAACCATCAAGGAGACTTTAACTAAAT 6253
1017 euThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVal 1033
6254 TAACGCTTGCAACTGGCTCTAGAGACTGGGTGCTCCTACTCCCCTTAGCC 6303
1034 LeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGl 1050
6304 CTGTACCGAGCCCGCAACACGCCGGGCCCCCATGGCCTCACCCCATATGA 6353

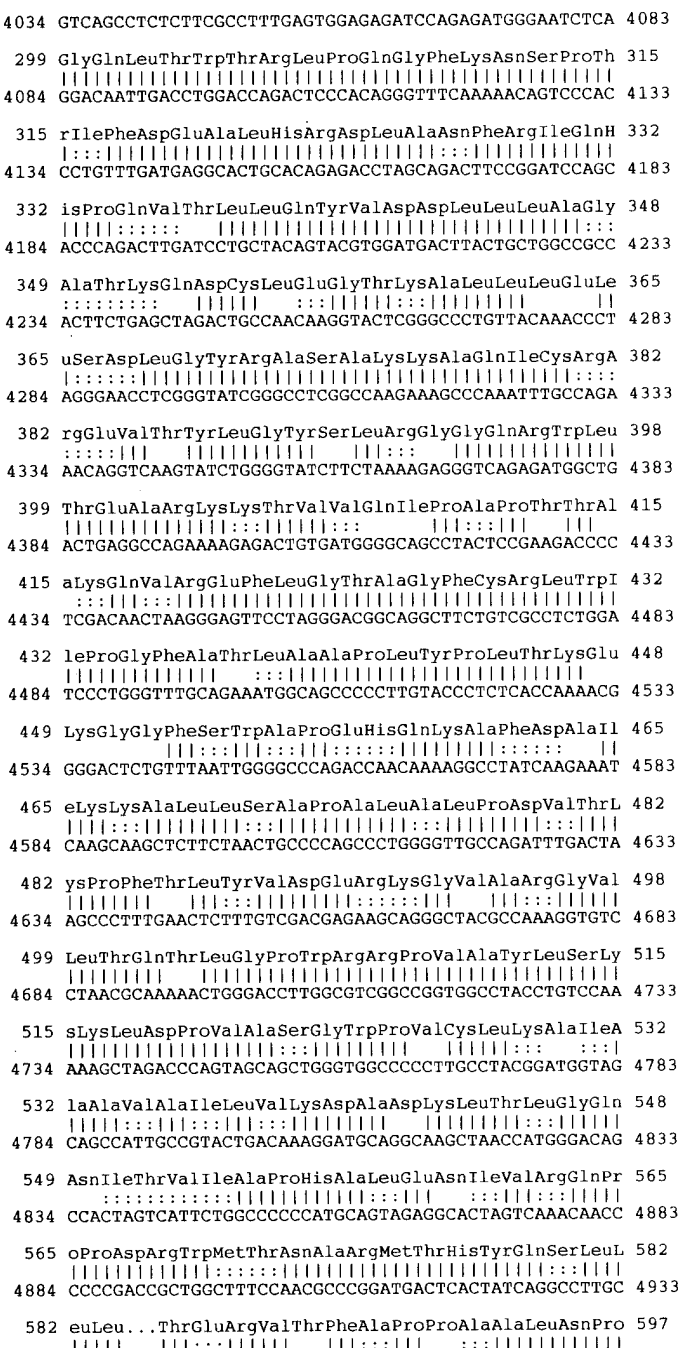
```
 1050 uLeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisS 1067
      |:::|||||||||:::|||||||||||||:::::::::
 6354 GATCTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGACCCTGACA 6403

 1067 erAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeu 1083
        :::   |||      |||      |||   ::::::|||:::||||||
 6404 TGACAAGAGTTACTAACAGCCCCTCTCTCCAAGCTCACTTACAGGCTCTC 6453

 1084 GluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGl 1100
            |||:::::::::::::|||:::   |||      ||||||:::
 6454 TACTTAGTCCAGCACGAAGTCTGGAGACCTCTGGCGGCAGCCTACCAA.. 6501

 1100 yGlyAspLeuGln.........ValProHisArgPheGlnValGlyAspS 1114
        :::|||:::         |||||||||   ::::::|||||||||:
 6502 .GAACAACTGGACCGACCGGTGGTACCTCACCCTTACCGAGTCGGCGACA 6550

 1114 erValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLys 1130
      ::|||:::|||||||||||||::::::   |||||||||   |||||||||
 6551 CAGTGTGGGTCCGCCGACACCAGACTAAGAACCTAGAACCTCGCTGGAAA 6600

 1131 GlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGl 1147
      |||||||||   |||||||||||||||||||||||:::|||||||:::||
 6601 GGACCTTACACAGTCCTGCTGACCACCCCCACCGCCCTCAAAGTAGACGG 6650

 1147 yIleProThrTrpIleHisAlaSerHisValLysProAlaProPro.... 1162
      ||||   :::|||||||||||||:::|||||||||   |||   |||
 6651 CATCGCAGCTTGGATACACGCCGCCCACGTGAAGGCTGCCGACCCCGGGG 6700

 1163 .....ProAspSer......GlyTrpArgAlaGluLysThrGluAsnPro 1175
           |||:::|||         ||||||:::::::::::::::||||||
 6701 GTGGACCATCCTCTAGACTGACATGGCGCGTTCAACGCTCTCAAAACCCC 6750

 1176 LeuLysLeuArgLeuHisArgLeuValPro 1185
      ||||||:::|||||||   |||   :::|||
 6751 TTAAAAATAAGGTTAACCCGCGAGGCCCCC 6780
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-4

```
seq_documentation_block:
; Sequence 4, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 7308
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
US-09-011-745-4


alignment_scores:
            Quality: 3967.00            Length:   1210
              Ratio:  3.959              Gaps:     13
 Percent Similarity:  82.810   Percent Identity: 62.479

alignment_block:
 US-09-171-553B-5 x US-09-011-745-4   ..
```

Align seg 1/1 to: US-09-011-745-4 from: 1 to: 7308

```
    1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
      |||   :::|||:::::|||   |||||||||||:::|||||||||||
 3184 GGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGT 3233

   17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerV 34
      |   ||||||||||||   |||||||||||||||||||||||:::|||||||
 3234 CGGGGGGCAACCCGTCACCTTCCTGGTAGATACTGGGGCCCAACACTCCG 3283

   34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
      |||||   |||      |||   |||::::::|||::::::|||||:::
 3284 TGCTGACCCAAAATCCTGGACCCCTAAGTGATAAGTCTGCCTGGGTCCAA 3333

   51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
      ||||||||||||   :::::|||   |||||||||   |||   |||
 3334 GGGGCTACTGGAGGAAAGCGGTATCGCTGGACCACGGATCGCAAAGTACA 3383

   67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
      |||:::::::|||:::|||||||||||||||||||   :::|||:::||||
 3384 TCTAGCTACCGGTAAGGTCACCCACTCTTTCCTCCATGTACCAGACTGTC 3433

   84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
      ||   |||||||||||||||||||||||||||||:::   |||||||||
 3434 CCTATCCTCTGTTAGGAAGAGATTTGCTGACTAAACTAAAAGCCCAAATC 3483

  101 SerPheGlu...GlnGlyArgProGluValSerValAsnAsnLysProIl 116
      ||||||   :::|||         ::::::      ::::::|||::
 3484 CACTTTGAGGGATCAGGAGCTCAGGTTATGGGACCAATGGGGCAGCCCCT 3533

  116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
      :   |||||||||||:::::::::|||||||:::|||||::::::   :
 3534 GCAAGTGTTGACCCTAAATATAGAAGATGAGCATCGGCTACATGAGACCT 3583

  133 lnValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
      ::   :::||||||   ::::::   |||   ||||||::::::||||||
 3584 CAAAAGAGCCAGATGTTTCTCTAGGGTCCACATGGCTGTCTGATTTTCCT 3633

  149 GlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPr 165
      |||||||||||||||||:::|||||||||||||||   :::      ||
 3634 CAGGCCTGGGCGGAAACCGGGGGCATGGGACTGGCAGTTCGCCAAGCTCC 3683

  165 oGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnT 182
      |   :::|||   |||||||||:::::|||||||||||:::::::||||
 3684 TCTGATCATACCTCTGAAAGCAACCTCTACCCCCGTGTCCATAAAACAAT 3733

  182 yrProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArg 198
      |||||:::|||:::||||||||   ||||||   |||||||:::||||||
 3734 ACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGA 3783

  199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPr 215
      |||::::::|||||||||||||||||||   ||||||||||||||||||||
 3784 CTGTTGGACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCC 3833

  215 oLeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnA 232
      ||||||||||||:::|||||||||||||||||||||||||||||||||||
 3834 CCTGCTACCCGTTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGG 3883

  232 spLeuArgGluValAsnLysArgValGlnAspIleHisProThrValPro 248
      ||||||||||||||||||||||||||:::|||||||||||||||||||||
 3884 ATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCC 3933

  249 AsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrTh 265
      ||||||||||||||||||||:::|||||||||:::::::::|||||||||
 3934 AACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACAC 3983

  265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 3984 TGTGCTTGATTTAAAGGATGCCTTTTTCTGCCTGAGACTCCACCCCACCA 4033

  282 erGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
      |||||||||||||||||||||||||||||||||      |||   :::
```


4034 GTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA 4083
299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
4084 GGACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCAC 4133
315 rIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
4134 CCTGTTTGATGAGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGC 4183
332 isProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGly 348
4184 ACCCAGACTTGATCCTGCTACAGTACGTGGATGACTTACTGCTGGCCGCC 4233
349 AlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLe 365
4234 ACTTCTGAGCTAGACTGCCAACAAGGTACTCGGGCCCTGTTACAAACCCT 4283
365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
4284 AGGGAACCTCGGGTATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGA 4333
382 rgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
4334 AACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTG 4383
399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
4384 ACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 4433
415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
4434 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGA 4483
432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
4484 TCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACG 4533
449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIl 465
4534 GGGACTCTGTTTAATTGGGGCCCAGACCAACAAAAGGCCTATCAAGAAAT 4583
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
4584 CAAGCAAGCTCTTCTAACTGCCCCAGCCCTGGGGTTGCCAGATTTGACTA 4633
482 ysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
4634 AGCCCTTTGAACTCTTTGTCGACGAGAAGCAGGGCTACGCCAAAGGTGTC 4683
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLy 515
4684 CTAACGCAAAAACTGGGACCTTGGCGTCGGCCGGTGGCCTACCTGTCCAA 4733
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
4734 AAAGCTAGACCCAGTAGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAG 4783
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
4784 CAGCCATTGCCGTACTGACAAAGGATGCAGGCAAGCTAACCATGGGACAG 4833
549 AsnIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPr 565
4834 CCACTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTCAAACAACC 4883
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
4884 CCCCGACCGCTGGCTTTCCAACGCCCGGATGACTCACTATCAGGCCTTGC 4933
582 euLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
4934 TTTTGGACACGGACCGGGTCCAGTTCGGACCGGTGGTAGCCCTGAACCCG 4983
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHi 614
4984 GCTACGCTGCTCCCA...CTGCCTGAGGAAGGGCTGCAACACAACTGCCT 5030
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
5031 TGATATCCTGGCCGAAGCCCACGGAACCCGACCCGACCTAACGGACCAGC 5080
631 roLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrVal 647
5081 CGCTCCCAGACGCCGACCACACCTGGTACACGGATGGAAGCAGTCTCTTA 5130
648 ValGluGlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgTh 664
5131 CAAGAGGGACAGCGTAAGGCGGGAGCTGCGGTGACCACCGAGACCGAGGT 5180
664 rIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluL 681
5181 AATCTGGGCTAAAGCCCTGCCAGCCGGGACATCCGCTCAGCGGGCTGAAC 5230
681 euMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
5231 TGATAGCACTCACCCAGGCCCTAAAGATGGCAGAAGGTAAGAAGCTAAAT 5280
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
5281 GTTTATACTGATAGCCGTTATGCTTTTGCTACTGCCCATATCCATGGAGA 5330
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
5331 AATATACAGAAGGCGTGGGTTGCTCACATCAGAAGGCAAAGAGATCAAAA 5380
731 snLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
5381 ATAAAGACGAGATCTTGGCCCTACTAAAAGCCCTCTTTCTGCCCAAAAGA 5430
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSe 764
5431 CTTAGCATAATCCATTGTCCAGGACATCAAAAGGGACACAGCGCCGAGGC 5480
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaV 781
5481 TAGAGGCAACCGGATGGCTGACCAAGCGGCCCGAAAGGCAGCC..... 5523
781 alAsnLeuLeuProIleIleGluThrProLysAla..... 792
5524 .....ATCACAGAGACTCCAGACACCTCTACCCTCCTCATA 5559
793 .....ProGluProArgArgGlnTyrThrLeuGluAs 803
5560 GAAAATTCATCACCCTACACCTCAGAACATTTTCATTACACAGTGACTGA 5609
803 pTrpGlnGluIleLysLysIleAspGln...PheSerGluThrProGluG 819
5610 TATAAAGGACCTAACCAAGTTGGGGGCCATTTATGATAAAACAAAGAAGT 5659
819 lyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGly 835
5660 ATTGGGTCTAC......CAAGGAAAACCTGTGATGCCTGACCAGTTTACT 5703
836 LeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHi 852
5704 TTTGAATTATTAGACTTTCTTCATCAGCTGACTCACCTCAGCTTCTCAAA 5753
852 sLeuGlnGlnLeuValArgThr......SerProTyrHisValLeuArgL 867
5754 AATGAAGGCTCTCCTAGAGAGAAGCCACAGTCCCTACTACATGCTGAACC 5803
867 euProGlyValAlaAspSerValValLysHisCysValProCysGlnLeu 883
5804 GGGATCGAACACTCAAAAATATCACTGAGACCTGCAAAGCTTGTGCACAA 5853

```
 884 ValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySe 900
5854 GTCAACGCCAGCAAGTCTGCCGTTAAACAGGGAACTAGGGTCCGCGGGCA 5903

 900 rHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaL 917
5904 TCGGCCCGGCACTCATTGGGAGATCGATTTCACCGAGATAAAGCCCGGAT 5953

 917 ysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrp 933
5954 TGTATGGCTATAAATATCTTCTAGTTTTTATAGATACCTTTTCTGGCTGG 6003

 934 ValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLy 950
6004 ATAGAAGCCTTCCCAACCAAGAAAGAAACCGCCAAGGTCGTAACCAAGAA 6053

 950 sIleLeuGluGluIleProArgPheGlyIleProLysValIleGlyS 967
6054 GCTACTAGAGGAGATCTTCCCCAGGTTCGGCATGCCTCAGGTATTGGGAA 6103

 967 erAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLys 983
6104 CTGACAATGGGCCTGCCTTCGTCTCCAAGGTGAGTCAGACAGTGGCCGAT 6153

 984 IleLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSe 1000
6154 CTGTTGGGGATTGATTGGAAATTACATTGTGCATACAGACCCCAAAGCTC 6203

1000 rGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysL 1017
6204 AGGCCAGGTAGAAAGAATGAATAGAACCATCAAGGAGACTTTAACTAAAT 6253

1017 euThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVal 1033
6254 TAACGCTTGCAACTGGCTCTAGAGACTGGGTGCTCCTACTCCCCTTAGCC 6303

1034 LeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGl 1050
6304 CTGTACCGAGCCCGCAACACGCCGGGCCCCCATGGCCTCACCCCATATGA 6353

1050 uLeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisS 1067
6354 GATCTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGACCCTGACA 6403

1067 erAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeu 1083
6404 TGACAAGAGTTACTAACAGCCCCTCTCTCCAAGCTCACTTACAGGCTCTC 6453

1084 GluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGl 1100
6454 TACTTAGTCCAGCACGAAGTCTGGAGACCTCTGGCGGCAGCCTACCAA.. 6501

1100 yGlyAspLeuGln.........ValProHisArgPheGlnValGlyAspS 1114
6502 .GAACAACTGGACCGACCGGTGGTACCTCACCCTTACCGAGTCGGCGACA 6550

1114 erValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLys 1130
6551 CAGTGTGGGTCCGCCGACACCAGACTAAGAACCTAGAACCTCGCTGGAAA 6600

1131 GlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGl 1147
6601 GGACCTTACACAGTCCTGCTGACCACCCCCACCGCCCTCAAAGTAGACGG 6650

1147 yIleProThrTrpIleHisAlaSerHisValLysProAlaProPro.... 1162
6651 CATCGCAGCTTGGATACACGCCGCCCACGTGAAGGCTGCCGACCCCGGGG 6700

1163 .....ProAspSer......GlyTrpArgAlaGluLysThrGluAsnPro 1175
6701 GTGGACCATCCTCTAGACTGACATGGCGCGTTCAACGCTCTCAAAACCCC 6750

1176 LeuLysLeuArgLeuHisArgLeuValPro 1185
```

```
6751 TTAAAAATAAGGTTAACCCGCGAGGCCCCC 6780
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-2

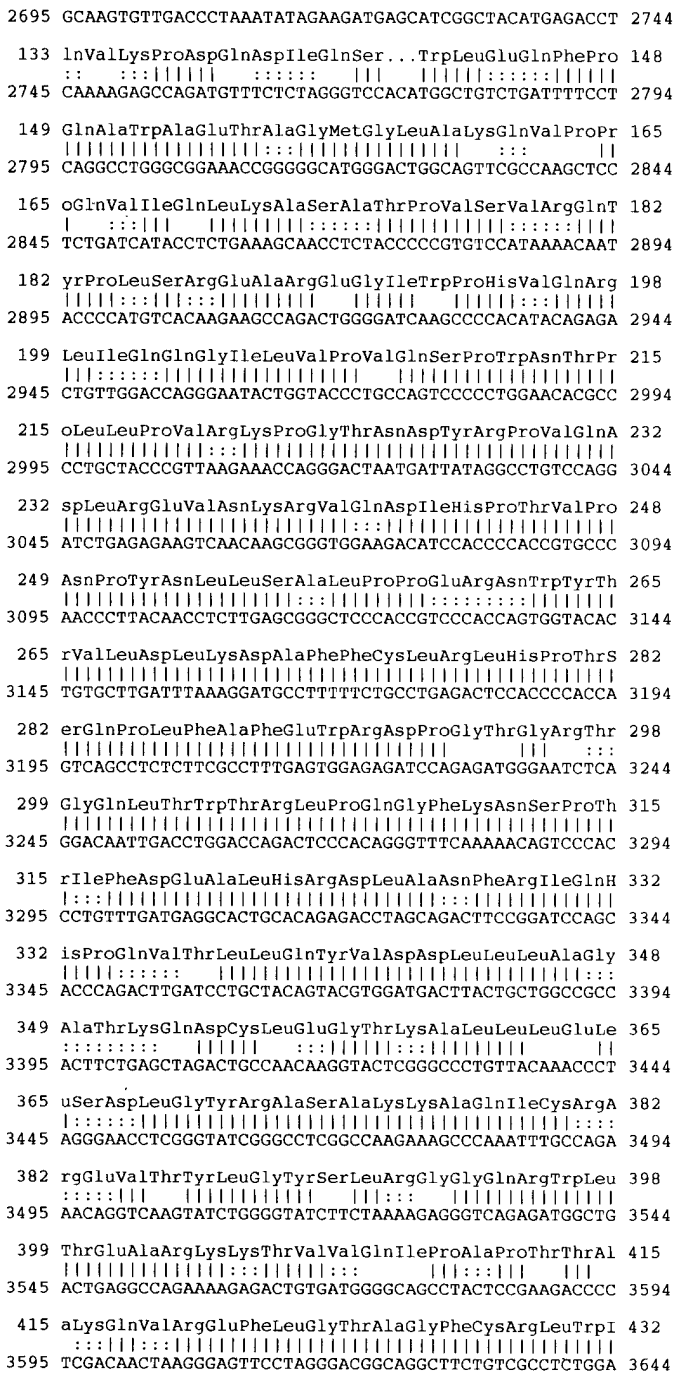
seq_documentation_block:
; Sequence 2, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 7616
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
US-09-011-745-2

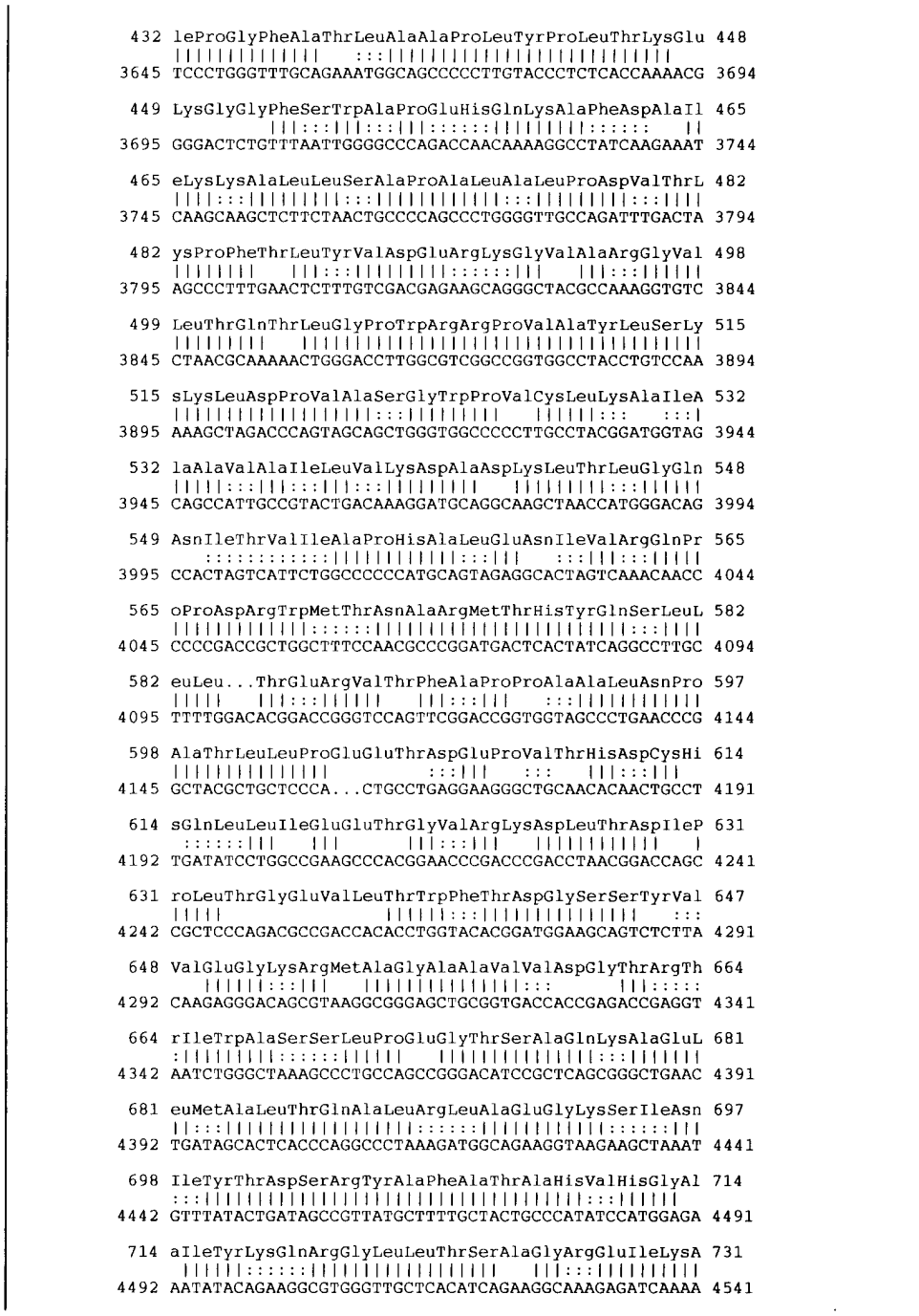
alignment_scores:
Quality: 3967.00 Length: 1210
Ratio: 3.959 Gaps: 13
Percent Similarity: 82.810 Percent Identity: 62.479

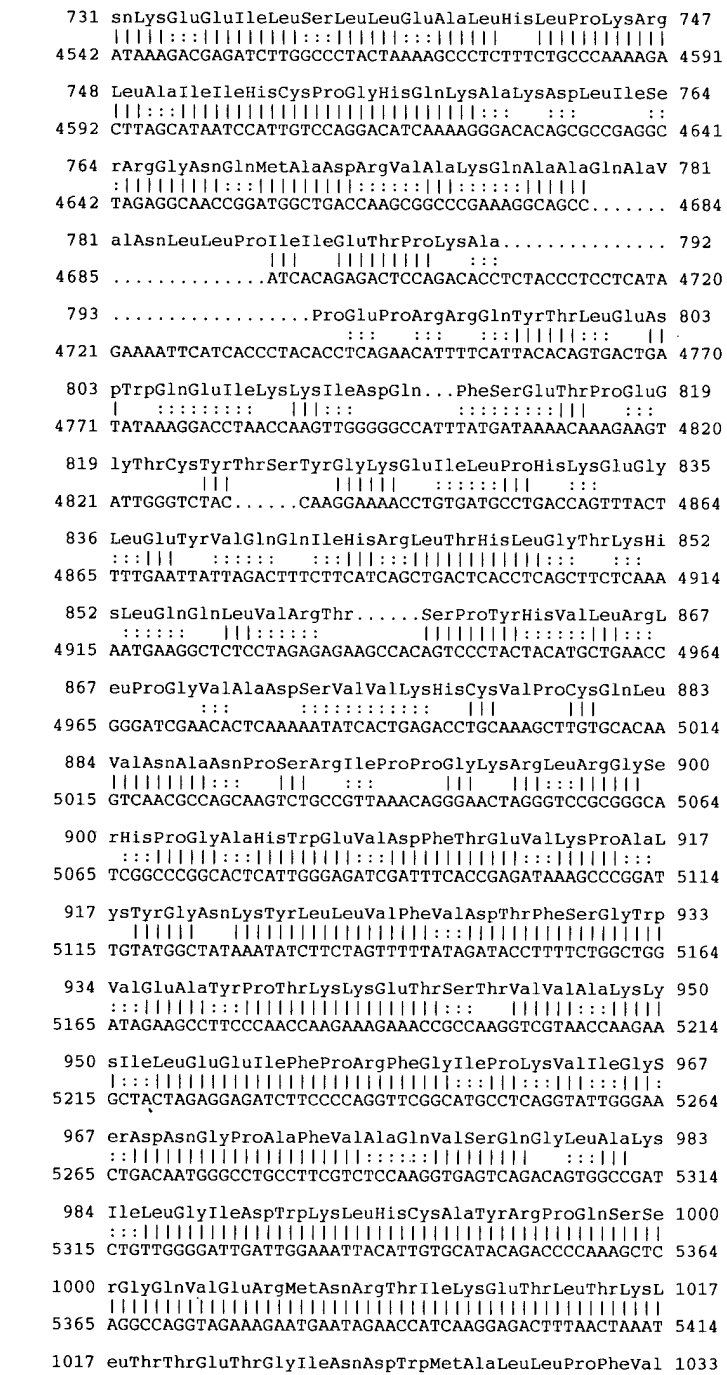
alignment_block:
US-09-171-553B-5 x US-09-011-745-2 ..

Align seg 1/1 to: US-09-011-745-2 from: 1 to: 7616

```
   1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVa 17
2345 GGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGT 2394

  17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerV 34
2395 CGGGGGGCAACCCGTCACCTTCCTGGTAGATACTGGGGCCCAACACTCCG 2444

  34 alLeuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
2445 TGCTGACCCAAAATCCTGGACCCCTAAGTGATAAGTCTGCCTGGGTCCAA 2494

  51 GlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAs 67
2495 GGGGCTACTGGAGGAAAGCGGTATCGCTGGACCACGGATCGCAAAGTACA 2544

  67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
2545 TCTAGCTACCGGTAAGGTCACCCACTCTTTCCTCCATGTACCAGACTGTC 2594

  84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
2595 CCTATCCTCTGTTAGGAAGAGATTTGCTGACTAAACTAAAAGCCCAAATC 2644

 101 SerPheGlu...GlnGlyArgProGluValSerValAsnAsnLysProIl 116
2645 CACTTTGAGGGATCAGGAGCTCAGGTTATGGGACCAATGGGGCAGCCCCT 2694

 116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
```


2695 GCAAGTGTTGACCCTAAATATAGAAGATGAGCATCGGCTACATGAGACCT 2744
133 lnValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
2745 CAAAAGAGCCAGATGTTTCTCTAGGGTCCACATGGCTGTCTGATTTTCCT 2794
149 GlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPr 165
2795 CAGGCCTGGGCGGAAACCGGGGGCATGGGACTGGCAGTTCGCCAAGCTCC 2844
165 oGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnT 182
2845 TCTGATCATACCTCTGAAAGCAACCTCTACCCCCGTGTCCATAAAACAAT 2894
182 yrProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArg 198
2895 ACCCCATGTCACAAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGA 2944
199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPr 215
2945 CTGTTGGACCAGGGAATACTGGTACCCTGCCAGTCCCCCTGGAACACGCC 2994
215 oLeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnA 232
2995 CCTGCTACCCGTTAAGAAACCAGGGACTAATGATTATAGGCCTGTCCAGG 3044
232 spLeuArgGluValAsnLysArgValGlnAspIleHisProThrValPro 248
3045 ATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCC 3094
249 AsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrTh 265
3095 AACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACAC 3144
265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
3145 TGTGCTTGATTTAAAGGATGCCTTTTTCTGCCTGAGACTCCACCCCACCA 3194
282 erGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
3195 GTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA 3244
299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
3245 GGACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCAC 3294
315 rIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
3295 CCTGTTTGATGAGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGC 3344
332 isProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGly 348
3345 ACCCAGACTTGATCCTGCTACAGTACGTGGATGACTTACTGCTGGCCGCC 3394
349 AlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLe 365
3395 ACTTCTGAGCTAGACTGCCAACAAGGTACTCGGGCCCTGTTACAAACCCT 3444
365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
3445 AGGGAACCTCGGGTATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGA 3494
382 rgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
3495 AACAGGTCAAGTATCTGGGGTATCTTCTAAAAGAGGGTCAGAGATGGCTG 3544
399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
3545 ACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTACTCCGAAGACCCC 3594
415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
3595 TCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCCTCTGGA 3644


432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
3645 TCCCTGGGTTTGCAGAAATGGCAGCCCCCTTGTACCCTCTCACCAAAACG 3694
449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIl 465
3695 GGGACTCTGTTTAATTGGGGCCCAGACCAACAAAAGGCCTATCAAGAAAT 3744
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
3745 CAAGCAAGCTCTTCTAACTGCCCCAGCCCTGGGGTTGCCAGATTTGACTA 3794
482 ysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
3795 AGCCCTTTGAACTCTTTGTCGACGAGAAGCAGGGCTACGCCAAAGGTGTC 3844
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLy 515
3845 CTAACGCAAAAACTGGGACCTTGGCGTCGGCCGGTGGCCTACCTGTCCAA 3894
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
3895 AAAGCTAGACCCAGTAGCAGCTGGGTGGCCCCCTTGCCTACGGATGGTAG 3944
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
3945 CAGCCATTGCCGTACTGACAAAGGATGCAGGCAAGCTAACCATGGGACAG 3994
549 AsnIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPr 565
3995 CCACTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTCAAACAACC 4044
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
4045 CCCCGACCGCTGGCTTTCCAACGCCCGGATGACTCACTATCAGGCCTTGC 4094
582 euLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
4095 TTTTGGACACGGACCGGGTCCAGTTCGGACCGGTGGTAGCCCTGAACCCG 4144
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHi 614
4145 GCTACGCTGCTCCCA...CTGCCTGAGGAAGGGCTGCAACACAACTGCCT 4191
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
4192 TGATATCCTGGCCGAAGCCCACGGAACCCGACCCGACCTAACGGACCAGC 4241
631 roLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrVal 647
4242 CGCTCCCAGACGCCGACCACACCTGGTACACGGATGGAAGCAGTCTCTTA 4291
648 ValGluGlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgTh 664
4292 CAAGAGGGACAGCGTAAGGCGGGAGCTGCGGTGACCACCGAGACCGAGGT 4341
664 rIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluL 681
4342 AATCTGGGCTAAAGCCCTGCCAGCCGGGACATCCGCTCAGCGGGCTGAAC 4391
681 euMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
4392 TGATAGCACTCACCCAGGCCCTAAAGATGGCAGAAGGTAAGAAGCTAAAT 4441
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
4442 GTTTATACTGATAGCCGTTATGCTTTTGCTACTGCCCATATCCATGGAGA 4491
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
4492 AATATACAGAAGGCGTGGGTTGCTCACATCAGAAGGCAAAGAGATCAAAA 4541


731 snLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
4542 ATAAAGACGAGATCTTGGCCCTACTAAAAGCCCTCTTTCTGCCCAAAAGA 4591
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSe 764
4592 CTTAGCATAATCCATTGTCCAGGACATCAAAAGGGACACAGCGCCGAGGC 4641
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaV 781
4642 TAGAGGCAACCGGATGGCTGACCAAGCGGCCCGAAAGGCAGCC..... 4684
781 alAsnLeuLeuProIleIleGluThrProLysAla..... 792
4685 .....ATCACAGAGACTCCAGACACCTCTACCCTCCTCATA 4720
793 .....ProGluProArgArgGlnTyrThrLeuGluAs 803
4721 GAAAATTCATCACCCTACACCTCAGAACATTTTCATTACACAGTGACTGA 4770
803 pTrpGlnGluIleLysLysIleAspGln...PheSerGluThrProGluG 819
4771 TATAAAGGACCTAACCAAGTTGGGGGCCATTTATGATAAAACAAAGAAGT 4820
819 lyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGly 835
4821 ATTGGGTCTAC......CAAGGAAAACCTGTGATGCCTGACCAGTTTACT 4864
836 LeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHi 852
4865 TTTGAATTATTAGACTTTCTTCATCAGCTGACTCACCTCAGCTTCTCAAA 4914
852 sLeuGlnGlnLeuValArgThr......SerProTyrHisValLeuArgL 867
4915 AATGAAGGCTCTCCTAGAGAGAAGCCACAGTCCCTACTACATGCTGAACC 4964
867 euProGlyValAlaAspSerValValLysHisCysValProCysGlnLeu 883
4965 GGGATCGAACACTCAAAAATATCACTGAGACCTGCAAAGCTTGTGCACAA 5014
884 ValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySe 900
5015 GTCAACGCCAGCAAGTCTGCCGTTAAACAGGGAACTAGGGTCCGCGGGCA 5064
900 rHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaL 917
5065 TCGGCCCGGCACTCATTGGGAGATCGATTTCACCGAGATAAAGCCCGGAT 5114
917 ysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrp 933
5115 TGTATGGCTATAAATATCTTCTAGTTTTTATAGATACCTTTTCTGGCTGG 5164
934 ValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLy 950
5165 ATAGAAGCCTTCCCAACCAAGAAAGAAACCGCCAAGGTCGTAACCAAGAA 5214
950 sIleLeuGluGluIleProArgPheGlyIleProLysValIleGlyS 967
5215 GCTACTAGAGGAGATCTTCCCCAGGTTCGGCATGCCTCAGGTATTGGGAA 5264
967 erAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLys 983
5265 CTGACAATGGGCCTGCCTTCGTCTCCAAGGTGAGTCAGACAGTGGCCGAT 5314
984 IleLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSe 1000
5315 CTGTTGGGGATTGATTGGAAATTACATTGTGCATACAGACCCCAAAGCTC 5364
1000 rGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysL 1017
5365 AGGCCAGGTAGAAAGAATGAATAGAACCATCAAGGAGACTTTAACTAAAT 5414
1017 euThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVal 1033
5415 TAACGCTTGCAACTGGCTCTAGAGACTGGGTGCTCCTACTCCCCTTAGCC 5464
1034 LeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGl 1050
5465 CTGTACCGAGCCCGCAACACGCCGGGCCCCCATGGCCTCACCCCATATGA 5514
1050 uLeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisS 1067
5515 GATCTTATATGGGGCACCCCCGCCCCTTGTAAACTTCCCTGACCCTGACA 5564
1067 erAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeu 1083
5565 TGACAAGAGTTACTAACAGCCCCTCTCTCCAAGCTCACTTACAGGCTCTC 5614
1084 GluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGl 1100
5615 TACTTAGTCCAGCACGAAGTCTGGAGACCTCTGGCGGCAGCCTACCAA.. 5662
1100 yGlyAspLeuGln.....ValProHisArgPheGlnValGlyAspS 1114
5663 .GAACAACTGGACCGACCGGTGGTACCTCACCCTTACCGAGTCGGCGACA 5711
1114 erValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLys 1130
5712 CAGTGTGGGTCCGCCGACACCAGACTAAGAACCTAGAACCTCGCTGGAAA 5761
1131 GlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGl 1147
5762 GGACCTTACACAGTCCTGCTGACCACCCCCACCGCCCTCAAAGTAGACGG 5811
1147 yIleProThrTrpIleHisAlaSerHisValLysProAlaProPro.... 1162
5812 CATCGCAGCTTGGATACACGCCGCCCACGTGAAGGCTGCCGACCCCGGGG 5861
1163 .....ProAspSer......GlyTrpArgAlaGluLysThrGluAsnPro 1175
5862 GTGGACCATCCTCTAGACTGACATGGCGCGTTCAACGCTCTCAAAACCCC 5911
1176 LeuLysLeuArgLeuHisArgLeuValPro 1185
5912 TTAAAAATAAGGTTAACCCGCGAGGCCCCC 5941

OM of: US-09-171-553B-6 to: Issued_Patents_NA:* out_format : pfs

Date: Feb 24, 2002 10:18 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09171553/runat_22022002_122826_11171/app_query.fasta_1.3312
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=feb22std.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09171553_@CGN1_1_158 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-171-553B-6
Query length: 656
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 134.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation	..		
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-3	+	3433.00	5859.62	0.0	3482	!				
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-1	+	2594.50	4421.86	1.2e-238	2462	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-2	+	2497.50	4241.08	1.4e-228	7333	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-1	+	2360.00	4003.38	2.4e-215	8060	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-3	+	2353.00	3991.22	1.2e-214	8132	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-25	+	2136.00	3635.91	7.1e-195	2000	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-30	+	2136.00	3635.91	7.1e-195	2000	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-24	+	1627.00	2764.19	2.6e-146	1493	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-7	+	1508.50	2541.95	6.1e-134	6312	!				
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-1	+	1486.50	2500.26	1.3e-131	8535	!				
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-075-272-1	+	1431.00	2404.62	2.7e-126	8655	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-8	+	1299.00	2178.09	1.1e-113	8323	!				
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-8	+	1299.00	2178.09	1.1e-113	8323	!				
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-8	+	1299.00	2178.09	1.1e-113	8323					
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-6	+	1258.00	2111.62	5.7e-110	6061	!				
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-850-961-1	+	1258.00	2107.55	9.7e-110	8332	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-7	+	1235.50	2087.65	1.2e-108	1914	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-5	+	1231.50	2066.11	2.0e-107	6028	!				
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-850-961-13	+	1227.00	2072.46	8.7e-108	2001	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-9	+	1226.00	2070.97	1.1e-107	1965	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-13	-	1208.00	2021.76	5.8e-105	8202	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-9	+	1204.50	2025.15	3.8e-105	3925	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-8	+	1191.50	2011.99	2.0e-104	1911	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-552-369-19	+	1173.50	1980.58	1.1e-102	1979	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-105-483-310	+	1163.50	1960.40	1.5e-101	2499	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-709-209-310	+	1163.50	1960.40	1.5e-101	2499	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-101-310	+	1163.50	1960.40	1.5e-101	2499	!				
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-5	+	1024.50	1702.43	3.6e-87	10970	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-18	+	640.00	1076.30	2.7e-52	698	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-073-836-1	+	634.00	1040.55	2.6e-50	5109	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-073-836-2	+	634.00	1040.38	2.7e-50	5176	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-278-630A-7	+	626.50	1044.12	1.7e-50	1407	!				
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-484-126-2	+	626.50	1043.77	1.7e-50	1446	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-278-630A-8	+	605.50	1008.73	1.5e-48	1329	!				
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-278-630A-9	+	601.00	1000.57	4.4e-48	1374	!				
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-484-126-4	+	600.00	999.02	5.4e-48	1356	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-10	+	484.00	819.75	5.2e-38	278	!				
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-4	+	433.00	685.87	1.5e-30	10306	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-8	+	430.00	726.96	7.7e-33	276	!	S			
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-1	+	423.50	687.53	1.2e-30	2518	!				
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-8	+	423.50	676.73	4.8e-30	5865					
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-007-282B-1	+	423.00	691.66	7.1e-31	170					
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-3	+	420.50	665.19	2.1e-29	966					
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-7	+	396.00	662.38	3.0e-29	445					
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-691-563C-58	+	295.50	472.23	1.2e-18	17					

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-3

seq_documentation_block:
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

alignment_scores:
Quality: 3433.00 Length: 658
Ratio: 5.257 Gaps: 2
Percent Similarity: 99.240 Percent Identity: 98.176

alignment_block:
US-09-171-553B-6 x US-09-111-085-3 ..

Align seg 1/1 to: US-09-111-085-3 from: 1 to: 3482

```
       1 MetHisProThrLeuSerArgArgHisLeuProThrArgGlyGlyGluPr 17
         |||||||||||||||||||   ||||||||||||||||||||||||||||||
     911 ATGCATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGAGAGCC 960

      17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     961 GAAAAGACTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTA 1010

      34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1011 CTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTATAGACAGCTCG 1060

      51 AsnProHisArgProLeuSerLeuThrTrpLeuIleIleAspProAspTh 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1061 AACCCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGATAC 1110

      67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1111 GGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGT 1160

      84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1161 GGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGC 1210

     101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyTh 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1211 ACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCAC 1260

     117 rGluLysGluLysTyrCysGlyGlySerGlyGluSerPheCysArgArgT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1261 AGAGAAAGAGAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGAT 1310

     134 rpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSerLeu 150
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1311 GGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGCCGATCTCTCTC 1360

 151 GlnAspArgValLysPheSerPheValAsnSerGlyProGlyLysTyrLy 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1361 CAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACAA 1410

 167 sMetMetLysLeuTyrLysAspLysSerCysSerProSerAspLeuAspT 184
     |:::||||||||||||||||||||||||||||||||||||||||||||||
1411 AGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATT 1460

 184 yrLeuLysIleSerPheThrGlu..ArgLysThrGlyLysTyrSerLysV 200
     ||||||||||||||||||||||  ||||||||||||||||||||||||||
1461 ATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAAG 1510

 200 alAspLysTrpTyrGluLeuGlyAsnSerPheLeu.LeuTyrGlyGlyGl 216
     ||||||||||||||||||||||||||||||||||    ||||||||||||
1511 TGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGG 1560

 216 yAlaGlySerThrLeuThrIleArgLeuArgIleGluThrGlyThrGluP 233
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1561 AGCAGGGTCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAAC 1610

 233 roProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyProPro 249
     ||||||||||||:::||||||||||||||||||||||||||||||||||||
1611 CCCCTGTGGCAGTGGGACCCGATAAAGTACTGGCTGAACAGGGGCCCCCG 1660

 250 AlaLeuGluProProHisAsnLeuProValProGlnLeuThrSerLeuAr 266
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1661 GCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGCG 1710

 266 gProAspIleThrGlnProProSerAsnSerThrThrGlyLeuIleProT 283
     ||||||||||||||||||||||||||||||:::||||||||||||||||||
1711 GCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGGATTGATTCCTA 1760

 283 hrAsnThrProArgAsnSerProGlyValProValLysThrGlyGlnArg 299
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1761 CCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGA 1810

 300 LeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSerThrAspPr 316
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1811 CTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCC 1860

 316 oAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProProTyrT 333
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1861 TGATGCCACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATT 1910

 333 yrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGluHisArg 349
     ||||||||||||||||||  ||||||||||||||||||||||||||||||
1911 ATGAGGGGATGGCTAAAGAAGGAAAATTCAATGTGACCAAAGAGCATAGA 1960

 350 AsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThrGluValSe 366
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1961 AATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTTC 2010

 366 rGlyLysGlyThrCysIleGlyLysAlaProProSerHisGlnHisLeuC 383
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2011 CGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTT 2060

 383 ysTyrSerThrValValTyrGluGlnAlaSerGluAsnGlnTyrLeuVal 399
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2061 GCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTA 2110

 400 ProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLeuThrProCysVa 416
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2111 CCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGT 2160

 416 lSerThrSerValPheAsnGlnSerLysAspPheCysValMetValGlnI 433
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2161 TTCCACCTCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAA 2210

 433 leValProArgValTyrTyrHisProGluGluValValLeuAspGluTyr 449
     ||||||||||||||||||||||||||||||||||||||||||||||||||

     2211 TCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCCTTGATGAATAT 2260

      450 AspTyrArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAl 466
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2261 GACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAGC 2310

      466 aValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGlyThrAlaA 483
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2311 TGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTG 2360

      483 laLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHis 499
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2361 CCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACAT 2410

      500 AlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAsnLe 516
          ||||||||||||||||||||||||||||||:::||||||||||||||||||
     2411 GCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCT 2460

      516 uGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgG 533
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2461 AGAAGAGTCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGG 2510

      533 lyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLys 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2511 GATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAA 2560

      550 GluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMe 566
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2561 GAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCAT 2610

      566 tAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArgGluAlaA 583
          |:::||||||||||:::::||||||||||||||||||||||||||||||
     2611 GAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTG 2660

      583 spGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMetThrThr 599
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2661 ACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACCACC 2710

      600 LeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuTh 616
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2711 CTGCTTTCTGCTCTGACGGGACCCCTAGTAGTCCTGCTCCTGTTACTTAC 2760

      616 rValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgV 633
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2761 AGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAG 2810

      633 alSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeuLeu 649
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2811 TGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTG 2860

      650 SerGlnGlyGluThrAspLeu 656
          |||||||||||||||||||||
     2861 AGCCAAGGAGAAACTGACCTC 2881

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-1

seq_documentation_block:
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
;  APPLICANT: Stoye, Jonathan P
;  APPLICANT: Weiss, Robin A
;  TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
;  TITLE OF INVENTION: specific sequences
;  FILE REFERENCE: 4238/75168
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2462
```

```
;   TYPE: DNA
;   ORGANISM: Porcine retrovirus
US-09-111-085-1


alignment_scores:
            Quality: 2594.50              Length:    680
              Ratio:  4.420                Gaps:     11
 Percent Similarity:  86.324   Percent Identity: 73.088

alignment_block:
 US-09-171-553B-6 x US-09-111-085-1   ..

 Align seg 1/1  to: US-09-111-085-1  from: 1  to: 2462

       1 MetHisProThrLeuSerArgArgHisLeuProThrArgGlyGlyGluPr 17
         |||||||||||||||||||||||||||||||||   ||||||||||:::||
     211 ATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCGGGGTGGAAAGCC 260

      17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
         ||||||||||:::|||||||||||||||||||||||||||||||||||||
     261 GAAAAGACTGAAAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTA 310

      34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
         |||||:::|||||||||||:::::::::|||||||||:::||||||
     311 CTCTGTCAATAACTCCTCAAGTTAATGGTAAACGCCTTGTGGACAGCCCG 360

      51 AsnProHisArgProLeuSerLeuThrTrpLeuIleIleAspProAspTh 67
         |||   |||:::|||||||||||||||||||:::   |||      ||
     361 AACTCCCATAAACCCTTATCTCTCACCTGGTTACTTACTGACTCCGGTAC 410

      67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
         ||||:::::::::|||||||||:::|||   ||||||   ||||||||||
     411 AGGTATTAATATTAACAGCACTCAAGGGGAGGCTCCCTTGGGGACCTGGT 460

      84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaVal..... 98
         ||||||||||:::   |||||||||   :::   |||::::::
     461 GGCCTGAATTATATGTCTGCCTTCGATCAGTAATCCCTGGTCTCAATGAC 510

      99 LysSerThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysPr 115
         ::::::|||||||||:::::::::|||:::|||||||||||||   |||||
     511 CAGGCCACACCCCCCGATGTACTCCGTGCTTACGGGTTTTACGTTTGCCC 560

     115 oGly...ThrGluLysGluLysTyrCysGlyGlySerGlyGluSerPheC 131
         ||||      :::::::|||:::||||||||||:::      :::   ||||
     561 AGGACCCCCAAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTT 610

     131 ysArgArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpPro 147
         ||:::::|||||||||:::|||||||||||||||:::|||||||||||||
     611 GCAAGCAATGGAGCTGCATAACTTCTAATGATGGGAATTGGAAATGGCCA 660

     148 IleSerLeuGlnAspArgValLysPheSerPheValAsnSer........ 161
         :::|||   ||||||||||||:::::|||||||||||||:::
     661 GTCTCTCAGCAAGACAGAGTAAGTTACTCTTTTGTTAACAATCCTACCAG 710

     162 ...................GlyProGlyLysTyrLys......MetMetL 170
                             |||   |||:::::|||      :::::::
     711 TTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGATTGGCAACAGC 760

     170 ysLeuTyrLysAspLys..............SerCysSerProSerAsp 181
         :::::   ||||||                |||||          |||
     761 GGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTCATTCGTTAGAC 810

     182 LeuAspTyrLeuLysIleSerPheThrGlu..ArgLysThrGlyLysTyr 197
         |||||||||||||||||||||||||||||  |||||||||   ||||||
     811 CTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATAT 860

     198 SerLysValAspLysTrpTyrGluLeuGlyAsnSer.PheLeuLeuTyrG 214
         |||||||||   |||||||||   ||||||||||| :::      ||||
     861 TCAAAAGTGGGTAAATGGTATATCTTGGGGAATAGTGTAC......TATG 904

     214 lyGlyGlyAla.........GlySerThrLeuThrIleArgLeuArgIle 227
         |||||::::::         ||||||:::|||||||||||||||||||||
```

```
     905 GAGGCTCTGGGAGAAAGAAAGGATCTGTTCTGACTATTCGCCTCAGAATA 954

     228 GluThrGlyThrGluProProValAlaMetGlyProAspLysValLeuAl 244
         ||||||      |||||||||||||||:::|||||||:::|||   |||||
     955 GAAACTCAGATGGAACCTCCGGTTGCTATAGGACCAAATAAGGGTTTGGC 1004

     244 aGluGlnGlyProProAlaLeuGluProProHisAsnLeuProValProG 261
         ||||||||||||||||      |||
    1005 CGAACAAGGACCTCCAATCCAAGAACAG..................... 1032

     261 lnLeuThrSerLeuArgProAspIleThrGlnProProSerAsnSerThr 277
                       |||||||:::   :::             |||:::|||
    1033 .............AGGCCATCTCCTAACCCCTCTGATTACAATACAACC 1068

     278 ThrGlyLeuIleProThrAsnThrProArgAsnSerProGlyValProVa 294
         :::|||   :::||||||               :::|||::::::   ::
    1069 TCTGGATCAGTCCCCACT..............GAGCCTAACATCACTAT 1103

     294 lLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaI 311
         :|||||||||   :::||||||||||||||||||||||||||||||||||:
    1104 TAAAACAGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTC 1153

     311 leAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSer 327
         ::|||||||||   |||:::|||||||||||||||||||||||||||:::
    1154 TTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCT 1203

     328 SerGlyProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnVa 344
         ||||||||||||||||||||||||||||||::      ||||||||||
    1204 TCGGGCCCACCTTACTATGAGGGAATGGCTAGAGGAGGGAAATTCAATGT 1253

     344 lThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuT 361
         ||||||||||||||||:::|||||||||||||||||:::||||||||||
    1254 GACAAAGGAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTA 1303

     361 hrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProPro 377
         ||||||||||||||||||||||||||||||||||   :::||||||
    1304 CCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGATGGTTCCCCCA 1353

     378 SerHisGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGl 394
         ||||||||||||||||      |||   ::::::::::::::||||||
    1354 TCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGA 1403

     394 uAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrG 411
         |:::|||||||||||||||||||:::|||||||||||||||||||||||
    1404 GAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTG 1453

     411 lyLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPhe 427
         ||||||||||||||||||||||   ||||||||||||:::||||||||||
    1454 GATTAACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAAAGACTTT 1503

     428 CysValMetValGlnIleValProArgValTyrTyrHisProGluGluVa 444
         |||||||||||||||||||||||||||||||||||:::||||||:::::
    1504 TGCGTTATGGTCCAAATTGTCCCCCGGGTGTACTACTATCCCGAAAAAGC 1553

     444 lValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProV 461
         :||||||||||||||||||||||||||||||||||||||||||||||||:
    1554 AGTCCTTGATGAATATGACTATAGATATAATCGGCCAAAAAGAGAGCCCA 1603

     461 alSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyVal 477
         ::|||||||||||||||||||||||||||||||||:::|||:::||||||
    1604 TATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGGCGTG 1653

     478 GlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGl 494
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1654 GGAACAGGAACGGCTGCCCTAATCACAGGACCGCAACAGCTGGAGAAAGG 1703

     494 yLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysG 511
         ||||:::::::||||||      :::||||||||||||:::|||||||::::
    1704 ACTTAGTAACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAA 1753

     511 luSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValVal 527
         ::|||||||||||||||||||||||||||||||||||||||||||||||||
    1754 AATCTGTCAGTAACCTGGAGGAATCCCTAACCTCCTTATCTGAAGTGGTT 1803
```

```
      528 LeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLe 544
          ||||||||||||||||||||||||||||||||||||||||:::|||||||||||
     1804 CTACAGAACAGAAGGGGGTTAGATCTGTTATTTCTAAAAGAAGGAGGGTT 1853

      544 uCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyA 561
          ||||:::||||||||||||||||||||||||||||||||||||||||||||
     1854 ATGTGTAGCCTTAAAAGAGGAATGCTGCTTCTATGTAGATCACTCAGGAG 1903

      561 laIleArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArg 577
          ||||||||||||||||:::|||||||||:::::|||||||||||||||||
     1904 CCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGA 1953

      578 ArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSe 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     1954 AGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTC 2003

      594 rProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValL 611
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     2004 TCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCC 2053

      611 euLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAla 627
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     2054 TGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCC 2103

      628 PheValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGl 644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     2104 TTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACA 2153

      644 nTyrGlnGlyLeuLeuSerGlnGlyGluThrAspLeu 656
          |||||||||||||||||||||||||||||||||||||
     2154 GTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTC 2190

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-2

seq_documentation_block:
; Sequence 2, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-2


alignment_scores:
            Quality: 2497.50              Length:    678
              Ratio:  4.328                Gaps:     13
 Percent Similarity:  85.103   Percent Identity: 71.829

alignment_block:
 US-09-171-553B-6 x US-08-766-528-2   ..

 Align seg 1/1  to: US-08-766-528-2  from: 1  to: 7333

        4 ThrLeuSerArgArgHisLeuProThrArgGlyGlyGluProLysArgLe 20
          ||||||||||||           |||:::|||   |||   :::
     4757 ACCTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGA..... 4801

       20 uArg.IleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThr 36
            ||| ||||||||||||||||||||||||||||||||||||||||||:::
     4802 .AGGAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTCA 4850

       37 IleThrProGlnAlaSerSerLysArgLeuIleAspSerSerAsnProHi 53
          ||||||||||||:::::::::|||||||||:::||||||   |||   ||
     4851 ATAACTCCTCAAGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCA 4900

       53 sArgProLeuSerLeuThrTrpLeuIleIleAspProAspThrGlyValT 70
          |:::|||||||||||||||||||:::   |||       ||||||::::
     4901 TAAACCCTTATCTCTCACCTGGTTACTTACTGACTCCGGTACAGGTATTA 4950

       70 hrValAsnSerThrArgGlyValAlaProArgGlyThrTrpTrpProGlu 86
          :::::|||||||||:::|||   ||||||   ||||||||||||||||||
     4951 ATATTAACAGCACTCAAGGGGAGGCTCCCTTGGGGACCTGGTGGCCTGAA 5000

       87 LeuHisPheCysLeuArgLeuIleAsnProAlaVal......LysSerTh 101
          |||:::   |||||||||   :::   |||::::::      ::::::||
     5001 TTATATGTCTGCCTTCGATCAGTAATCCCTGGTCTCAATGACCAGGCCAC 5050

      101 rProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGly...T 117
          |||||||:::::::::|||:::||||||||||||   |||||||||
     5051 ACCCCCCGATGTACTCCGTGCTTACGGGTTTTACGTTTGCCCAGGACCCC 5100

      117 hrGluLysGluLysTyrCysGlyGlySerGlyGluSerPheCysArgArg 133
            ::::::|||:::|||||||||:::      :::   ||||||::::::
     5101 CAAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAA 5150

      134 TrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSerLe 150
          |||||||||:::|||||||||||||||:::||||||||||||:::|||
     5151 TGGAGCTGCATAACTTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCA 5200

      150 uGlnAspArgValLysPheSerPheValAsnSer.................. 161
            ||||||||||||:::::|||||||||||:::
     5201 GCAAGACAGAGTAAGTTACTCTTTTGTTAACAATCCTACCAGTTATAATC 5250

      162 ...........GlyProGlyLysTyrLys......MetMetLysLeuTyr 172
                     |||   |||::::::|||      ::::::::::::
     5251 AATTTAATTATGGCCATGGGAGATGGAAAGATTGGCAACAGCGGGTACAA 5300

      173 LysAspLys..............SerCysSerProSerAspLeuAspTy 184
          ||||||                ||||||         ||||||||||
     5301 AAAGATGTACGAAATAAGCAAATAAGCTGTCATTCGTTAGACCTAGATTA 5350

      184 rLeuLysIleSerPheThrGlu..ArgLysThrGlyLysTyrSerLysVa 200
          ||||||||||||||||||||||  |||||||||   ||||||||||||||
     5351 CTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGT 5400

      200 lAspLysTrpTyrGluLeuGlyAsnSer.PheLeuLeuTyrGlyGlyGly 216
          |   |||||||||   |||||||||||| :::      |||||||||:::
     5401 GGGTAAATGGTATATCTTGGGGAATAGTGTAC......TATGGAGGCTCT 5444

      217 Ala.........GlySerThrLeuThrIleArgLeuArgIleGluThrGl 230
          :::         ||||||:::|||||||||||||||||||||||||||||
```

```
5445 GGGAGAAAGAAAGGATCTGTTCTGACTATTCGCCTCAGAATAGAAACTCA 5494

 230 yThrGluProProValAlaMetGlyProAspLysValLeuAlaGluGlnG 247
5495 GATGGAACCTCCGGTTGCTATAGGACCAAATAAGGGTTTGGCCGAACAAG 5544

 247 lyProProAlaLeuGluProProHisAsnLeuProValProGlnLeuThr 263
5545 GACCTCCAATCCAAGAACAG..............................  5564

 264 SerLeuArgProAspIleThrGlnProProSerAsnSerThrThrGlyLe 280
5565 ......AGGCCATCTCCTAACCCCTCTGATTACAATACAACCTCTGGATC 5608

 280 uIleProThrAsnThrProArgAsnSerProGlyValProValLysThrG 297
5609 AGTCCCCACT..............GAGCCTAACATCACTATTAAAACAG 5643

 297 lyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSer 313
5644 GGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCC 5693

 314 ThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyPr 330
5694 ACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCC 5743

 330 oProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysG 347
5744 ACCTTACTATGAGGGAATGGCTAGAGGAGGGAAATTCAATGTGACAAAGG 5793

 347 luHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThr 363
5794 AACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTACCCTTACT 5843

 364 GluValSerGlyLysGlyThrCysIleGlyLysAlaProProSerHisGl 380
5844 GAGGTTTCTGGAAAAGGCACCTGCATAGGGATGGTTCCCCCATCCCACCA 5893

 380 nHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsnGlnT 397
5894 ACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCAAT 5943

 397 yrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLeuThr 413
5944 ATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACC 5993

 414 ProCysValSerThrSerValPheAsnGlnSerLysAspPheCysValMe 430
5994 CCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAAAGACTTTTGCGTTAT 6043

 430 tValGlnIleValProArgValTyrTyrHisProGluGluValValLeuA 447
6044 GGTCCAAATTGTCCCCCGGGTGTACTACTATCCCGAAAAAGCAGTCCTTG 6093

 447 spGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSerLeu 463
6094 ATGAATATGACTATAGATATAATCGGCCAAAAAGAGAGCCCATATCCCTG 6143

 464 ThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGl 480
6144 ACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGGCGTGGGAACAGG 6193

 480 yThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyG 497
6194 AACGGCTGCCCTAATCACAGGACCGCAACAGCTGGAGAAAGGACTTAGTA 6243

 497 luLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerVal 513
6244 ACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTC 6293

 514 SerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAs 530
6294 AGTAACCTGGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAA 6343

 530 nArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaA 547
6344 CAGAAGGGGGTTAGATCTGTTATTTCTAAAAGAAGGAGGGTTATGTGTAG 6393

 547 laLeuLysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArg 563
6394 CCTTAAAAGAGGAATGCTGCTTCTATGTAGATCACTCAGGAGCCATCAGA 6443

 564 AspSerMetAsnLysLeuArgLysLysLeuGluArgArgArgArgGluAr 580
6444 GACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGAAAG 6493

 580 gGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpM 597
6494 AGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGA 6543

 597 etThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeu 613
6544 TGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTGCTCCTG 6593

 614 LeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValAr 630
6594 TTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAG 6643

 630 gGluArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnG 647
6644 AGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAG 6693

 647 lyLeuLeuSerGlnGlyGluThrAspLeu 656
6694 GCCTTCTGAGCCAAGGAGAAACTGACCTC 6722
```

```
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-1

seq_documentation_block:
; Sequence 1, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1


alignment_scores:
            Quality: 2360.00              Length:    656
              Ratio:  4.252                Gaps:     11
 Percent Similarity:  84.604   Percent Identity: 68.445

alignment_block:
 US-09-171-553B-6 x US-08-766-528-1   ..

 Align seg 1/1  to: US-08-766-528-1  from: 1  to: 8060

       1 MetHisProThrLeuSerArgArgHisLeuProThrArgGlyGlyGluPr 17
         ||||||||||||||||||||||||||||||||   ||||||||||:::||
      86 ATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCGGGGTGGAAAGCC 135

      17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
         ||||||||||:::|||||||||||||||||||||||||||||||||||||
     136 GAAAAGACTGAAAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTA 185

      34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
         |||||:::|||||   |||:::::::::   |||:::   ||||||
     186 CTCTGTCAATAACCTCTCAGACTAATGGTATGCGCATAGGAGACAGCCTG 235

      51 AsnProHisArgProLeuSerLeuThrTrpLeuIleIleAspProAspTh 67
         |||   |||:::|||||||||||||||||||||||   |||      ||
     236 AACTCCCATAAACCCTTATCTCTCACCTGGTTAATTACTGACTCCGGCAC 285

      67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
         ||||:::::::::|||:::|||:::|||   ||||||   ||||||||||
     286 AGGTATTAATATCAACAACACTCAAGGGGAGGCTCCTTTAGGAACCTGGT 335

      84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
         |||||:::|||:::   |||||||||   :::   |||:::::   |||
     336 GGCCTGATCTATACGTTTGCCTCAGATCAGTTATTCCTAGTCTGACCTCA 385

     101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyTh 117
            ||||||:::::::::::::::::::|||||||||   |||||||||
     386 ...CCCCCAGATATCCTCCATGCTCACGGATTTTATGTTTGCCCAGGACC 432

     117 rGluLysGlu...LysTyrCysGlyGlySerGlyGluSerPheCysArgA 133
             ::::::   |||:::|||||:::      :::   ||||||::::
     433 ACCAAATAATGGAAAACATTGCGGAAATCCCAGAGATTTCTTTTGTAAAC 482

     133 rgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149
         ::|||:::|||||||||||||||||||||   ||||||||||||   |||
     483 AATGGAACTGTGTAACCTCTAATGATGGATATTGGAAATGGCCAACCTCT 532

     150 LeuGlnAspArgValLysPheSerPheValAsn......SerGlyProGl 164
            ||||||||||||:::|||||||::|||||      ::::::   ||
     533 CAGCAGGATAGGGTAAGTTTTTCTTATGTCAACACCTATACCAGCTCTGG 582

     164 yLysTyrLysMetMetLysLeuTyrLysAspLysSer......CysSerP 179
         |:::::::::   :::         :::      |||      |||||||
     583 ACAATTTAATTACCTGACCTGGATTAGAACTGGAAGCCCCAAGTGCTCTC 632

     179 roSerAspLeuAspTyrLeuLysIleSerPheThr..GluArgLysThrG 195
         ||||||||||||||||||||||||||||||||||||  ||||||||||||
     633 CTTCAGACCTAGATTACCTAAAAATAAGTTTCACTGAGAAAGGAAAACAA 682

     195 lyLysTyrSerLysValAspLysTrpTyrGluLeuGlyAsnSer.PheLe 211
           ||||||   |||:::   |||||||||   ||||||||||::: :::
     683 GAAAATATCCTAAAATGGGTAAATGGTATGTCTTGGGGAATGGTATAT.. 730

     211 uLeuTyrGlyGlyGlyAla.........GlySerThrLeuThrIleArgL 225
            |||||||||::::::         ||||||   ||||||||||||
     731 ....TATGGAGGCTCGGGTAAACAACCAGGCTCCATTCTAACTATTCGCC 776

     225 euArgIleGluThrGlyThrGluProProValAlaMetGlyProAspLys 241
         ||:::|||:::         ||||||||||:::|||:::||||||:::

     777 TCAAAATAAACCAG...CTGGAGCCTCCAATGGCTATAGGACCAAATACG 823

     242 ValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPr 258
         ||||||:::   |||   ||||||:::      |||
     824 GTCTTGACGGGTCAAAGACCCCCAACCCAAGGACCA.............. 859

     258 oValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerA 275
                                                |||   ||||
     860 ..................................GGACCATCCTCTA 872

     275 snSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 291
         ||   |||:::|||      ||||||::::::      ::::::
     873 ACATAACTTCTGGATCAGACCCCACTGAGTCTAGCAGCACGACT...... 916

     292 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPh 308
                 |||   |||   :::||||||||||||||||||||||||||||
     917 .........AAAATGGGGGCAAAACTTTTTAGCCTCATCCAGGGAGCTTT 957

     308 eGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLeuC 325
         |||||||:::|||||||||   |||:::||||||||||||||||||||||
     958 TCAAGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTAT 1007

     325 ysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluArgLys 341
         ||||||::   ||||||||||||||||||||||||||||:::::   |||
    1008 GCTTAGCTTTGGGCCCACCTTACTATGAAGGAATGGCTAGAAGAGGGAAA 1057

     342 PheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAs 358
         ||||||||||||||||||||||||||:::||||||||||||||||:::||
    1058 TTCAATGTGACAAAAGAACATAGAGACCAATGCACATGGGGATCCCAAAA 1107

     358 nLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysA 375
         ||||||||||||||||||||||||||||||||||||||||||||||||||:
    1108 TAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGAAAGG 1157

     375 laProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 391
         ::|||||||||||||||||||||||||      |||   :::::::::|||
    1158 TTCCCCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCAA 1207

     392 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCy 408
         :::|||||||:::|||||||||||||||||||:::|||||||||||||||
    1208 ACCTCTGAAAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATG 1257

     408 sAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSerL 425
         ||||||||||||||||||||||||||||||   ||||||||||||:::|
    1258 TAATACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTTAACCAAACTA 1307

     425 ysAspPheCysValMetValGlnIleValProArgValTyrTyrHisPro 441
         ||||||||||:::|||||||||||||||||||||||||||||||:::|||
    1308 AAGATTTTTGCATTATGGTCCAAATTGTTCCCCGAGTGTATTACTATCCC 1357

     442 GluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgProLysAr 458
         |||:::::::::|||||||||||||||||||   :::|||   |||||
    1358 GAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAAGAG 1407

     458 gGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaV 475
         |||||||:::|||||||||||||||||||||||||||||||||:::|||:
    1408 AGAACCCATATCTCTGACACTTGCTGTGATGCTCGGACTTGGAGTGGCAG 1457

     475 alGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 491
         ::|||||||||||||||||||||||||||:::|||||||||||||||||
    1458 CAGGTGTAGGAACAGGAACAGCTGCCCTGGTCACGGGACCACAGCAGCTA 1507

     492 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAl 508
         |||   ||||||::::::||||||      :::||||||||||||:::|||
    1508 GAAACAGGACTTAGTAACCTACATCGAATTGTAACAGAAGATCTCCAAGC 1557

     508 aLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSerG 525
         ||||::::::|||||||||||||||||||||||||||||||||||||||||
    1558 CCTAGAAAAATCTGTCAGTAACCTGGAGGAATCCCTAACCTCCTTATCTG 1607

     525 luValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGlu 541
         ||||||||||||||||||||||||||||||||||||||||||||:::|||
    1608 AAGTAGTCCTACAGAATAGAAGAGGGTTAGATTTATTATTTCTAAAAGAA 1657
```

```
     542 GlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHi 558
         ||||||||||||:::|||||||||||||||||||||||||||||||||||
    1658 GGAGGATTATGTGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTGGATCA 1707

     558 sSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGluA 575
         |||||||||||||||||||||||||||||||||||:::::|||||||:
    1708 TTCAGGGGCCATCAGAGACTCCATGAACAAACTTAGAGAAAGGTTGGAGA 1757

     575 rgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPhe 591
         ::||||||||||||:::|||:::   |||||||||||||||||||||||
    1758 AGCGTCGAAGGGAAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGGTTC 1807

     592 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLe 608
         ||||||||||||||:::::|||||||||||||||||||||||||||||||
    1808 AACAGGTCTCCTTGGTTGGCTACCCTACTTTCTGCTTTAACAGGACCCTT 1857

     608 uValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgP 625
         |:::|||||||||||||||||||||||||||||||:::|||||||::::
    1858 AATAGTCCTCCTCCTGTTACTCACAGTTGGGCCATGTATTATTAACAAGT 1907

     625 heValAlaPheValArgGluArgValSerAlaValGlnIleMetValLeu 641
         :::::||||||:::|||||||||:::|||||||||||||||||||||||
    1908 TAATTGCCTTCATTAGAGAACGAATAAGTGCAGTCCAGATCATGGTACTT 1957

     642 ArgGlnGlnTyrGln 646
         |||||||||||||||
    1958 AGACAACAGTACCAA 1972

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-3

seq_documentation_block:
; Sequence 3, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;     APPLICANT:  Jay A. Fishman
;     TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;     TITLE OF INVENTION:  AND METHODS OF USE
;     NUMBER OF SEQUENCES:  74
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;       STREET:  60 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109-1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/766,528
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/572,645
;       FILING DATE:  14-DEC-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Louis Myers
;       REGISTRATION NUMBER:  35,965
;       REFERENCE/DOCKET NUMBER:  MGP-038CP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227-5941
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8132 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-766-528-3

alignment_scores:
            Quality: 2353.00              Length:    656
              Ratio:  4.232                Gaps:     12
 Percent Similarity:  84.756   Percent Identity: 68.598

alignment_block:
 US-09-171-553B-6 x US-08-766-528-3   ..

 Align seg 1/1  to: US-08-766-528-3  from: 1  to: 8132

       1 MetHisProThrLeuSerArgArgHisLeuProThrArgGlyGlyGluPr 17
         ||||||||||||||||:::|||||||||||||||   |||||||||:::||
    5620 ATGCATCCCACGTTAAACCGGCGCCACCTCCCGATTCGGGGTGGAAAGCC 5669

      17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
         ||||||||||:::|||||||||||||||||||||||||||||||||||||
    5670 GAAAAGACTGAAAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTA 5719

      34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
         ||||:::||||||   |||:::::::::   |||:::   ||||||
    5720 CTCTGTCAATAACCTCTCAGACTAATGGTATGCGCATAGGAGACAGCCTG 5769

      51 AsnProHisArgProLeuSerLeuThrTrpLeuIleIleAspProAspTh 67
         |||   |||:::|||||||||||||||||||||||||   |||      ||
    5770 AACTCCCATAAACCCTTATCTCTCACCTGGTTAATTACTGACTCCGGCAC 5819

      67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
         ||||::::::::::|||:::|||:::|||   ||||||   ||||||||||
    5820 AGGTATTAATATCAACAACACTCAAGGGGAGGCTCCTTTAGGAACCTGGT 5869

      84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
         |||||:::|||:::   |||||||||   :::   |||::::::   |||
    5870 GGCCTGATCTATACGTTTGCCTCAGATCAGTTATTCCTAGTCTGACCTCA 5919

     101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyTh 117
            ||||||::::::::::::::::::|||||||||   |||||||||
    5920 ...CCCCCAGATATCCTCCATGCTCACGGATTTTATGTTTGCCCAGGACC 5966

     117 rGluLysGlu...LysTyrCysGlyGlySerGlyGluSerPheCysArgA 133
             ::::::   |||:::|||||:::      :::   ||||||::::
    5967 ACCAAATAATGGAAAACATTGCGGAAATCCCAGAGATTTCTTTTGTAAAC 6016

     133 rgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149
         ::|||:::|||||||||||||||||||||   ||||||||||||   |||
    6017 AATGGAACTGTGTAACCTCTAATGATGGATATTGGAAATGGCCAACCTCT 6066

     150 LeuGlnAspArgValLysPheSerPheValAsn......SerGlyProGl 164
            ||||||||||||:::||||||:::||||||      ::::::   ||
    6067 CAGCAGGATAGGGTAAGTTTTTCTTATGTCAACACCTATACCAGCTCTGG 6116

     164 yLysTyrLysMetMetLysLeuTyrLysAspLysSer......CysSerP 179
         |:::::::::   :::         :::       |||      |||||||
    6117 ACAATTTAATTACCTGACCTGGATTAGAACTGGAAGCCCCAAGTGCTCTC 6166

     179 roSerAspLeuAspTyrLeuLysIleSerPheThr..GluArgLysThrG 195
         ||||||||||||||||||||||||||||||||||   ||||||||||||
    6167 CTTCAGACCTAGATTACCTAAAAATAAGTTTCACTGAGAAAGGAAAACAA 6216

     195 lyLysTyrSerLysValAspLysTrpTyrGluLeuGlyAsnSer.PheLe 211
          ||||||   |||:::   |||||||||   |||||||||::: :::
    6217 GAAAATATCCTAAAATGGGTAAATGGTATGTCTTGGGGAATGGTATAT.. 6264

     211 uLeuTyrGlyGlyGlyAla.........GlySerThrLeuThrIleArgL 225
             ||||||||||:::::          ||||||   ||||||||||||
    6265 ....TATGGAGGCTCGGGTAAACAACCAGGCTCCATTCTAACTATTCGCC 6310

     225 euArgIleGluThrGlyThrGluProProValAlaMetGlyProAspLys 241
         ||:::|||:::         |||||||||:::|||:::|||||||:::
    6311 TCAAAATAAACCAG...CTGGAGCCTCCAATGGCTATAGGACCAAATACG 6357

     242 ValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPr 258
         ||||||:::   |||   ||||||:::      |||
```

```
6358 GTCTTGACGGGTCAAAGACCCCCAACCCAAGGACCA..............  6393

 258 oValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerA  275
                                          |||   ||||
6394 ..........................................GGACCATCCTCTA 6406

 275 snSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly  291
     ||   |||:::|||      ||||||:::::      ||||||
6407 ACATAACTTCTGGATCAGACCCCACTGAGTCT......AACAGC......  6444

 292 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPh  308
         :::|||   |||   :::||||||||||||||||||||||||||
6445 ...ACGACTAAAATGGGGGCAAAACTTTTTAGCCTCATCCAGGGAGCTTT 6491

 308 eGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLeuC  325
     |||||||:::|||||||||   |||:::||||||||||||||||||||||
6492 TCAAGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTAT  6541

 325 ysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluArgLys  341
     |||||:::|||||||||||||||||||||||||||||||:::::   |||
6542 GCTTAGCTTCGGGCCCACCTTACTATGAAGGAATGGCTAGAAGAGGGAAA 6591

 342 PheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAs  358
     ||||||||||||||||||||||:::|||||||||||||||||||:::||
6592 TTCAATGTGACAAAAGAACATAGAGACCAATGCACATGGGGATCCCAAAA 6641

 358 nLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysA  375
     ||||||||||||||||||||||||||||||||||||||||||||||||:
6642 TAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGAAAGG 6691

 375 laProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln  391
     ::|||||||||||||||||||||||      |||   :::::::::|||
6692 TTCCCCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCAA 6741

 392 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCy  408
     :::||||||:::|||||||||||||||||||||:::|||||||||||||
6742 ACCTCTGAGAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATG 6791

 408 sAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSerL  425
     |||||||||||||||||||||||||||||||   ||||||||||||:::|
6792 TAATACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTTAACCAAACTA 6841

 425 ysAspPheCysValMetValGlnIleValProArgValTyrTyrHisPro  441
     ||||||||||:::|||||||||||||||||||||||||||||||:::|||
6842 AAGATTTTTGCATTATGGTCCAAATTGTTCCCCGAGTGTATTACTATCCC 6891

 442 GluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgProLysAr  458
     |||:::::::::|||||||||||||||||||||   :::|||   |||||
6892 GAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAAGAG 6941

 458 gGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaV  475
     |||||||:::|||||||||||||||||||||||||||||||||||:::|||:
6942 AGAACCCATATCTCTGACACTTGCTGTGATGCTCGGACTTGGAGTGGCAG 6991

 475 alGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu  491
     ::|||||||||||||||||||||||||||:::|||||||||||||||||
6992 CAGGTGTAGGAACAGGAACAGCTGCCCTGGTCACGGGACCACAGCAGCTA 7041

 492 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAl  508
     |||   ||||||::::::|||||      :::|||||||||||||:::||
7042 GAAACAGGACTTAGTAACCTACATCGAATTGTAACAGAAGATCTCCAAGC 7091

 508 aLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSerG  525
     ||||::::::|||||||||||||||||||||||||||||||||||||||
7092 CCTAGAAAAATCTGTCAGTAACCTGGAGGAATCCCTAACCTCCTTATCTG 7141

 525 luValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGlu  541
     ||||||||||||||||||||||||||||||||||||||||||||:::|||
7142 AAGTAGTCCTACAGAATAGAAGAGGGTTAGATTTATTATTTCTAAAAGAA 7191

 542 GlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHi  558
     ||||||||||||:::||||||||||||||||||||||||||||||||||
7192 GGAGGATTATGTGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTGGATCA 7241

 558 sSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGluA  575
     ||||||||||||||||||||||||||||||||||||:::::::||||||:
7242 TTCAGGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAAGGTTGGAGA 7291

 575 rgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPhe  591
     ::|||||||||||||:::|||:::   ||||||||||||||||||||||||
7292 AGCGTCGAAGGGAAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGGTTC 7341

 592 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLe  608
     |||||||||   |||::::::||||||||||||||||||||||||||||
7342 AACAGGTCTCTTTGGTTGGCTACCCTACTTTCTGCTTTAACAGGACCCTT 7391

 608 uValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgP  625
     |:::|||||||||||||||||||||||||||||||:::|||||||::::
7392 AATAGTCCTCCTCCTGTTACTCACAGTTGGGCCATGTATTATTAACAAGT 7441

 625 heValAlaPheValArgGluArgValSerAlaValGlnIleMetValLeu  641
     :::::|||||||:::||||||||||:::||||||||||||||||||||||
7442 TAATTGCCTTCATTAGAGAACGAATAAGTGCAGTCCAGATCATGGTACTT 7491

 642 ArgGlnGlnTyrGln 646
     |||||||||||||||
7492 AGACAACAGTACCAA 7506
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-25

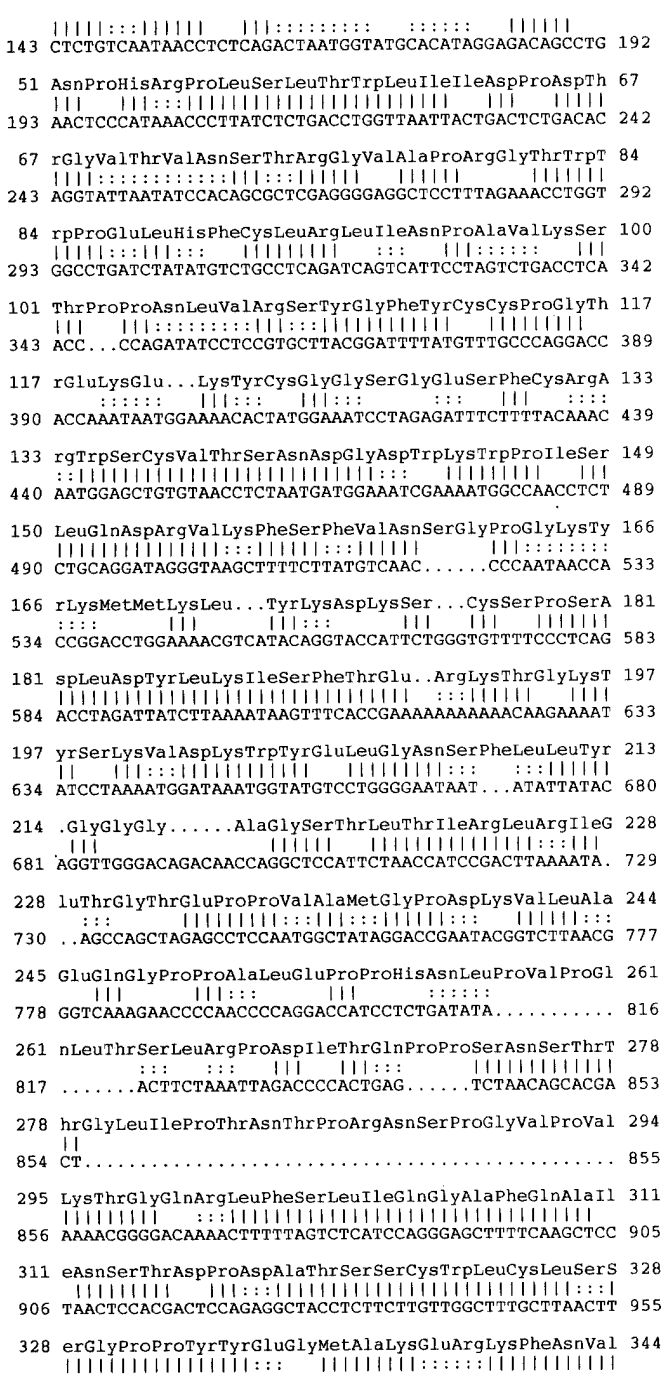
seq_documentation_block:
; Sequence 25, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Contiguous
; OTHER INFORMATION: compilation of SEQ ID NOs: 18, 22, and 24.
US-09-376-781-25

alignment_scores:
Quality: 2136.00 Length: 654
Ratio: 3.963 Gaps: 14
Percent Similarity: 82.416 Percent Identity: 64.526

alignment_block:
US-09-171-553B-6 x US-09-376-781-25 ..

Align seg 1/1 to: US-09-376-781-25 from: 1 to: 2000

```
   1 MetHisProThrLeuSerArgArgHisLeuProThrArgGlyGlyGluPr  17
     :::|||||||||||||||||||||||||||      :::|||   :::||
  43 ATACATCCCACGTTAAGCCGGCGCCACCTCCTGATTCAGGGTGAAAAGCC  92

  17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT  34
     ||||||||||:::|||||||||||||||||||||||||:::||||||||||
  93 AAAAAGACTAAAAATCCCCTTAAGCTTCGCCTCCATCACATGGTTCCTTA 142

  34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer  50
```

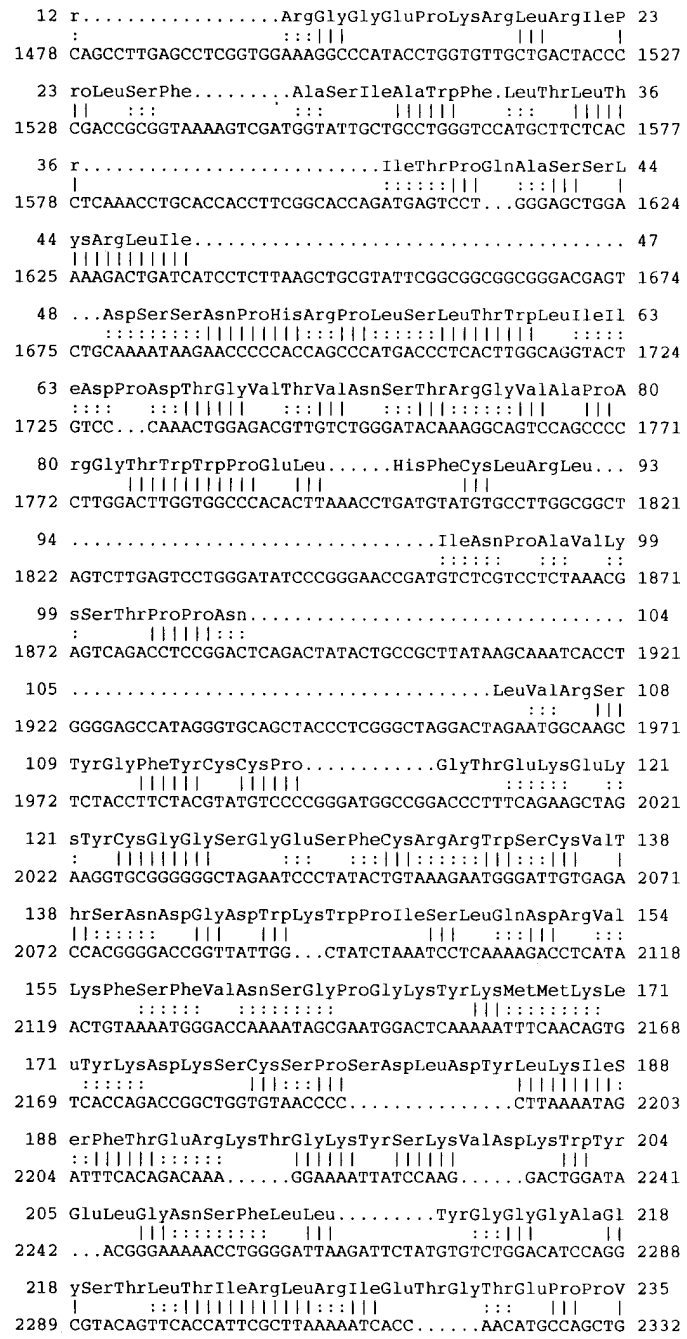

143 CTCTGTCAATAACCTCTCAGACTAATGGTATGCACATAGGAGACAGCCTG 192
51 AsnProHisArgProLeuSerLeuThrTrpLeuIleIleAspProAspTh 67
193 AACTCCCATAAACCCTTATCTCTGACCTGGTTAATTACTGACTCTGACAC 242
67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
243 AGGTATTAATATCCACAGCGCTCGAGGGGAGGCTCCTTTAGAAACCTGGT 292
84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
293 GGCCTGATCTATATGTCTGCCTCAGATCAGTCATTCCTAGTCTGACCTCA 342
101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyTh 117
343 ACC...CCAGATATCCTCCGTGCTTACGGATTTTATGTTTGCCCAGGACC 389
117 rGluLysGlu...LysTyrCysGlyGlySerGlyGluSerPheCysArgA 133
390 ACCAAATAATGGAAAACACTATGGAAATCCTAGAGATTTCTTTTACAAAC 439
133 rgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149
440 AATGGAGCTGTGTAACCTCTAATGATGGAAATCGAAAATGGCCAACCTCT 489
150 LeuGlnAspArgValLysPheSerPheValAsnSerGlyProGlyLysTy 166
490 CTGCAGGATAGGGTAAGCTTTTCTTATGTCAAC......CCCAATAACCA 533
166 rLysMetMetLysLeu...TyrLysAspLysSer...CysSerProSerA 181
534 CCGGACCTGGAAAACGTCATACAGGTACCATTCTGGGTGTTTTCCCTCAG 583
181 spLeuAspTyrLeuLysIleSerPheThrGlu..ArgLysThrGlyLysT 197
584 ACCTAGATTATCTTAAAATAAGTTTCACCGAAAAAAAAAAACAAGAAAAT 633
197 yrSerLysValAspLysTrpTyrGluLeuGlyAsnSerPheLeuLeuTyr 213
634 ATCCTAAAATGGATAAATGGTATGTCCTGGGGAATAAT...ATATTATAC 680
214 .GlyGlyGly......AlaGlySerThrLeuThrIleArgLeuArgIleG 228
681 AGGTTGGGACAGACAACCAGGCTCCATTCTAACCATCCGACTTAAAATA. 729
228 luThrGlyThrGluProProValAlaMetGlyProAspLysValLeuAla 244
730 ..AGCCAGCTAGAGCCTCCAATGGCTATAGGACCGAATACGGTCTTAACG 777
245 GluGlnGlyProProAlaLeuGluProProHisAsnLeuProValProGl 261
778 GGTCAAAGAACCCCAACCCCAGGACCATCCTCTGATATA..... 816
261 nLeuThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrT 278
817 .......ACTTCTAAATTAGACCCCACTGAG......TCTAACAGCACGA 853
278 hrGlyLeuIleProThrAsnThrProArgAsnSerProGlyValProVal 294
854 CT..... 855
295 LysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIl 311
856 AAAACGGGGACAAAACTTTTTAGTCTCATCCAGGGAGCTTTTCAAGCTCC 905
311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerS 328
906 TAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTTGCTTAACTT 955
328 erGlyProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnVal 344
956 CGGGCCCACCTTACTATAAGAAAATGGCTAAAAGAGAAAAATTCAATGTG 1005
345 ThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuTh 361
1006 ACAAAAAAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTAC 1055
361 rLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProProS 378
1056 CCTTACTGAGGTTTCTGGAAAGGACACCTGCATAAAAAAGGTTCCCCCAT 1105
378 erHisGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGlu 394
1106 CCCACCAACACCTTTACAACCACACTGAAGCCTTTAATCAAACCTCTGAG 1155
395 AsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGl 411
1156 AGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGG 1205
411 yLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheC 428
1206 ATTAACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAAAGACTTTT 1255
428 ysValMetValGlnIleValProArgValTyrTyrHisProGluGluVal 444
1256 ACATTATGGTCCAAATTGTTCCCCGAGTATATTACTATCCCAAGAAAACA 1305
445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProVa 461
1306 ATTCTCGATGAATATGATTACAGGAACCATCGACAAAAGAAAAAACCCAT 1355
461 lSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValG 478
1356 ATCCCTGACACTCGCAGTAATGCTCGGACTCGGAGTGATAACAGGTGTGA 1405
478 lyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGly 494
1406 GAACAGGAACTGCAGCTTTAGTTACAGGACCTCAGCAGCTAGAAACAGGA 1455
495 LeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGl 511
1456 CTTAGTAACCTACATCAAATTGTAACAGGAAATCTCCAAGCCCTAAAAAA 1505
511 uSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValL 528
1506 ATCTGTCAGTAACCTGGAAAAATCCCTAACCTCCTTATCTGAAGTAGTTC 1555
528 euGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeu 544
1556 TACAGAATAAAAAAGGGTTAGATTTATTATTTCTAAAAAAAAGAAGATTA 1605
545 CysAlaAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyAl 561
1606 TGTGTAGCCTTAAAGGAGAAATGCTGTTTTTATGTAGATCATTCAGGGGC 1655
561 aIleArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArgA 578
1656 CATCAGAGACTCCATGAACAAGCTTAAAAAAAGGTTGGAGAAACGTCGAA 1705
578 rgGluArgGluAla.AspGlnGlyTrpPheGluGlyTrpPheAsnArgSe 594
1706 GGGAAAAGGAAACTTACTCAAAGATGGTTTAAAAGATGGTTCAACAGGTC 1755
594 rProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValL 611
1756 TCCTTGGTTGGCTACCCTACTTTCTACTTTAACAGGACCCTTAATAGTCC 1805
611 euLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAla 627
1806 TCCTCCTGTTACTCACAGTTGGGCCATGTATTATTAACAAGTTAATTGCC 1855
628 PheValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGl 644
1856 TTCATTAGAAAACGAATAAGTGCAGTCCAGATCATGGTACTTAGACAACA 1905

```
     644 nTyrGln 646
         |||||||
    1906 GTACCAA 1912

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-30

seq_documentation_block:
; Sequence 30, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 2000
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: consensus sequence of PERV-D.
US-09-376-781-30


alignment_scores:
            Quality: 2136.00              Length:    654
              Ratio:  3.963                Gaps:     14
 Percent Similarity:  82.416   Percent Identity: 64.526

alignment_block:
 US-09-171-553B-6 x US-09-376-781-30   ..

 Align seq 1/1  to: US-09-376-781-30  from: 1  to: 2000

       1 MetHisProThrLeuSerArgArgHisLeuProThrArgGlyGlyGluPr 17
         :::|||||||||||||||||||||||||||||      :::|||   :::||
      43 ATACATCCCACGTTAAGCCGGCGCCACCTCCTGATTCAGGGTGAAAAGCC 92

      17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
         ||||||||||:::|||||||||||||||||||||||||:::||||||||||
      93 AAAAAGACTAAAAATCCCCTTAAGCTTCGCCTCCATCACATGGTTCCTTA 142

      34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
         |||||:::||||||   |||:::::::::   ::::::   ||||||
     143 CTCTGTCAATAACCTCTCAGACTAATGGTATGCACATAGGAGACAGCCTG 192

      51 AsnProHisArgProLeuSerLeuThrTrpLeuIleIleAspProAspTh 67
         |||   |||:::|||||||||||||||||||||||||   |||   |||||
     193 AACTCCCATAAACCCTTATCTCTGACCTGGTTAATTACTGACTCTGACAC 242

      67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
         ||||:::::::::::|||:::||||||   ||||||      |||||||
     243 AGGTATTAATATCCACAGCGCTCGAGGGGAGGCTCCTTTAGAAACCTGGT 292

      84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
         |||||:::|||:::   |||||||||   :::   |||:::::   |||
     293 GGCCTGATCTATATGTCTGCCTCAGATCAGTCATTCCTAGTCTGACCTCA 342

     101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyTh 117
         |||   |||:::::::::|||:::||||||||||||   |||||||||
     343 ACC...CCAGATATCCTCCGTGCTTACGGATTTTATGTTTGCCCAGGACC 389

     117 rGluLysGlu...LysTyrCysGlyGlySerGlyGluSerPheCysArgA 133

         ::::::   |||:::   |||:::      :::   |||   ::::
     390 ACCAAATAATGGAAAACACTATGGAAATCCTAGAGATTTCTTTTACAAAC 439

     133 rgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149
         ::|||||||||||||||||||||||||||||:::   |||||||||   |||
     440 AATGGAGCTGTGTAACCTCTAATGATGGAAATCGAAAATGGCCAACCTCT 489

     150 LeuGlnAspArgValLysPheSerPheValAsnSerGlyProGlyLysTy 166
         |||||||||||||||:::||||||:::||||||      |||::::::::
     490 CTGCAGGATAGGGTAAGCTTTTCTTATGTCAAC......CCCAATAACCA 533

     166 rLysMetMetLysLeu...TyrLysAspLysSer...CysSerProSerA 181
         ::::      |||      |||:::     |||   |||   |||||||
     534 CCGGACCTGGAAAACGTCATACAGGTACCATTCTGGGTGTTTTCCCTCAG 583

     181 spLeuAspTyrLeuLysIleSerPheThrGlu..ArgLysThrGlyLysT 197
         |||||||||||||||||||||||||||||||||  :::|||||   ||||
     584 ACCTAGATTATCTTAAAATAAGTTTCACCGAAAAAAAAAAACAAGAAAAT 633

     197 yrSerLysValAspLysTrpTyrGluLeuGlyAsnSerPheLeuLeuTyr 213
         ||   |||:::||||||||||||   |||||||||:::   :::||||||
     634 ATCCTAAAATGGATAAATGGTATGTCCTGGGGAATAAT...ATATTATAC 680

     214 .GlyGlyGly......AlaGlySerThrLeuThrIleArgLeuArgIleG 228
          |||            ||||||   ||||||||||||||||:::|||
     681 AGGTTGGGACAGACAACCAGGCTCCATTCTAACCATCCGACTTAAAATA. 729

     228 luThrGlyThrGluProProValAlaMetGlyProAspLysValLeuAla 244
           :::      |||||||||:::|||:::|||||||:::   |||||||:::
     730 ..AGCCAGCTAGAGCCTCCAATGGCTATAGGACCGAATACGGTCTTAACG 777

     245 GluGlnGlyProProAlaLeuGluProProHisAsnLeuProValProGl 261
           |||      |||:::      |||      ::::::
     778 GGTCAAAGAACCCCAACCCCAGGACCATCCTCTGATATA..... 816

     261 nLeuThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrT 278
             :::   :::   |||   |||:::      ||||||||||||
     817 .......ACTTCTAAATTAGACCCCACTGAG......TCTAACAGCACGA 853

     278 hrGlyLeuIleProThrAsnThrProArgAsnSerProGlyValProVal 294
         ||
     854 CT..................................................... 855

     295 LysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIl 311
         |||||||||   :::||||||||||||||||||||||||||||||||||||
     856 AAAACGGGGACAAAACTTTTTAGTCTCATCCAGGGAGCTTTTCAAGCTCC 905

     311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerS 328
         |||||||||   |||:::||||||||||||||||||||||||||||||:::|
     906 TAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTTGCTTAACTT 955

     328 erGlyProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnVal 344
         ||||||||||||||||:::   |||||||||:::::::||||||||||:::
     956 CGGGCCCACCTTACTATAAGAAAATGGCTAAAAGAGAAAAATTCAATGTG 1005

     345 ThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuTh 361
         ||||||:::||||||:::|||||||||||||||||:::||||||||||||
    1006 ACAAAAAAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTAC 1055

     361 rLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProProS 378
         ||||||||||||||||||||   |||||||||   |||:::|||||||
    1056 CCTTACTGAGGTTTCTGGAAAGGACACCTGCATAAAAAAGGTTCCCCCAT 1105

     378 erHisGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGlu 394
         ||||||||||||||          |||   :::::::::|||:::||||||
    1106 CCCACCAACACCTTTACAACCACACTGAAGCCTTTAATCAAACCTCTGAG 1155

     395 AsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGl 411
         :::|||||||||||||||||||:::||||||||||||||||||||||||||
    1156 AGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGG 1205

     411 yLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheC 428
         ||||||||||||||||||||||   ||||||||||||:::|||||||||
```

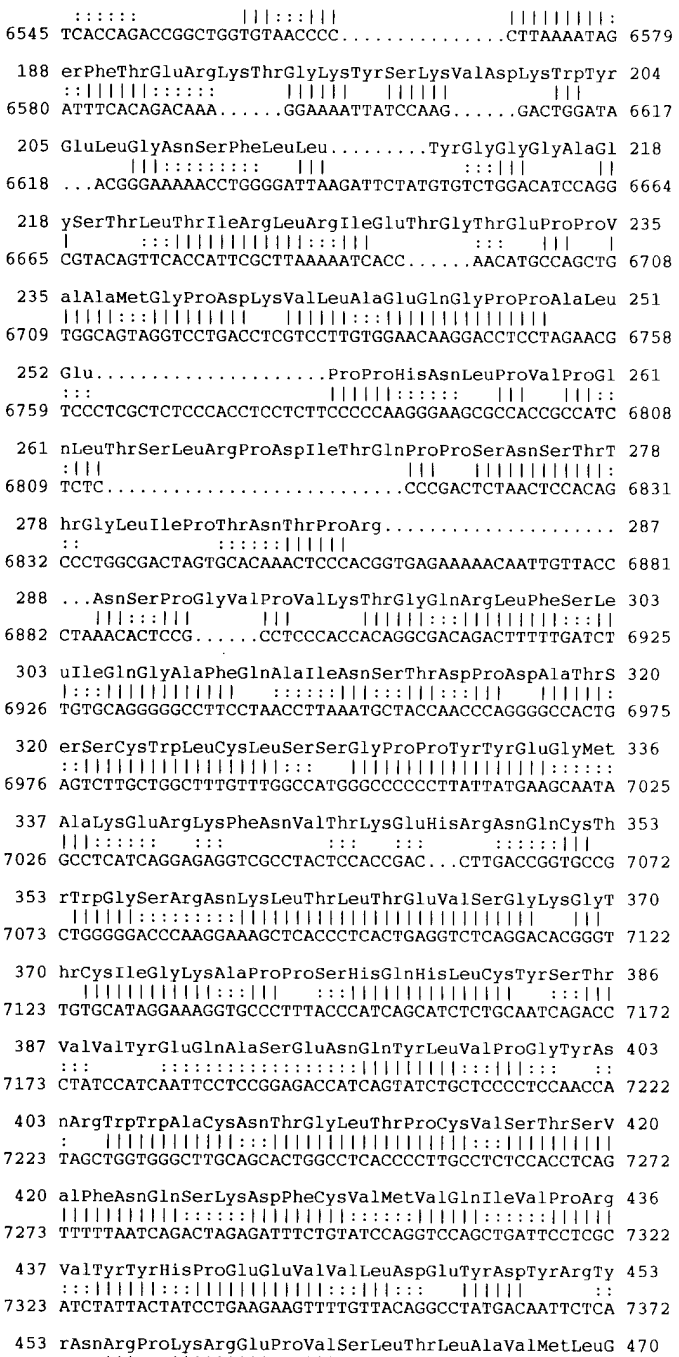
```
 1206 ATTAACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAAAGACTTTT 1255

  428 ysValMetValGlnIleValProArgValTyrTyrHisProGluGluVal 444
      :::||||||||||||||||||||||||||||||||:::|||:::::::::
 1256 ACATTATGGTCCAAATTGTTCCCCGAGTATATTACTATCCCAAGAAAACA 1305

  445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProVa 461
      :::|||||||||||||||||||||   :::|||   |||:::::::|||::
 1306 ATTCTCGATGAATATGATTACAGGAACCATCGACAAAAGAAAAAACCCAT 1355

  461 lSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValG 478
      :|||||||||||||||||||||||||||||||||:::   :::||||||
 1356 ATCCCTGACACTCGCAGTAATGCTCGGACTCGGAGTGATAACAGGTGTGA 1405

  478 lyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGly 494
       |||||||||||||||||:::|||||||||||||||||||||   |||
 1406 GAACAGGAACTGCAGCTTTAGTTACAGGACCTCAGCAGCTAGAAACAGGA 1455

  495 LeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGl 511
      |||:::::::||||||      :::|||   :::|||:::||||||||::
 1456 CTTAGTAACCTACATCAAATTGTAACAGGAAATCTCCAAGCCCTAAAAAA 1505

  511 uSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValL 528
      :|||||||||||||||||:::|||||||||||||||||||||||||||||
 1506 ATCTGTCAGTAACCTGGAAAAATCCCTAACCTCCTTATCTGAAGTAGTTC 1555

  528 euGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeu 544
      ||||||||::::::||||||||||||||||||||:::::      |||
 1556 TACAGAATAAAAAAGGGTTAGATTTATTATTTCTAAAAAAAAGAAGATTA 1605

  545 CysAlaAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyAl 561
      |||:::||||||||||||:::|||||||||||||||||||||||||||||
 1606 TGTGTAGCCTTAAAGGAGAAATGCTGTTTTTATGTAGATCATTCAGGGGC 1655

  561 aIleArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArgA 578
      ||||||||||||||||||||||||||:::|||:::|||||||:::|||||||
 1656 CATCAGAGACTCCATGAACAAGCTTAAAAAAAGGTTGGAGAAACGTCGAA 1705

  578 rgGluArgGluAla.AspGlnGlyTrpPheGluGlyTrpPheAsnArgSe 594
      |||||:::|||:::    |||   |||||||:::   ||||||||||||||
 1706 GGGAAAAGGAAACTTACTCAAAGATGGTTTAAAAGATGGTTCAACAGGTC 1755

  594 rProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValL 611
      |||||||::::::||||||||||||:::|||||||||||||||||:::|||||
 1756 TCCTTGGTTGGCTACCCTACTTTCTACTTTAACAGGACCCTTAATAGTCC 1805

  611 euLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAla 627
      ||||||||||||||||||||||||||||||||:::|||||||:::::::|||
 1806 TCCTCCTGTTACTCACAGTTGGGCCATGTATTATTAACAAGTTAATTGCC 1855

  628 PheValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGl 644
      |||:::|||:::|||:::||||||||||||||||||||||||||||||||
 1856 TTCATTAGAAAACGAATAAGTGCAGTCCAGATCATGGTACTTAGACAACA 1905

  644 nTyrGln 646
      |||||||
 1906 GTACCAA 1912

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-24

seq_documentation_block:
; Sequence 24, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
;   LENGTH: 1493
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: sequence of the 3' end of the PERV-D env region.
US-09-376-781-24


alignment_scores:
            Quality: 1627.00              Length:    491
              Ratio:  3.988                Gaps:     11
 Percent Similarity:  83.096   Percent Identity: 66.395

alignment_block:
 US-09-171-553B-6 x US-09-376-781-24   ..

 Align seg 1/1  to: US-09-376-781-24  from: 1  to: 1493

    163 ProGlyLysTyrLysMetMetLysLeu...TyrLysAspLysSer...Cy 177
        |||:::::::::::::      |||      |||:::      |||   ||
          4 CCCAATAACCACCGGACCTGGAAAACGTCATACAGGTACCATTCTGGGTG 53

    177 sSerProSerAspLeuAspTyrLeuLysIleSerPheThrGlu..ArgLy 193
        |   ||||||||||||||||||||||||||||||||||||||||  :::||
         54 TTTTCCCTCAGACCTAGATTATCTTAAAATAAGTTTCACCGAAAAAAAAA 103

    193 sThrGlyLysTyrSerLysValAspLysTrpTyrGluLeuGlyAsnSerP 210
        ||||   ||||||   |||:::||||||||||||   |||||||||:::
        104 AACAAGAAAATATCCTAAAATGGATAAATGGTATGTCCTGGGGAATAAT. 152

    210 heLeuLeuTyr.GlyGlyGly......AlaGlySerThrLeuThrIleAr 224
          :::|||||| |||               ||||||   ||||||||||
        153 ..ATATTATACAGGTTGGGACAGACAACCAGGCTCCATTCTAACCATCCG 200

    224 gLeuArgIleGluThrGlyThrGluProProValAlaMetGlyProAspL 241
        ||||:::|||   :::      |||||||||:::|||:::||||||:::
        201 ACTTAAAATA...AGCCAGCTAGAGCCTCCAATGGCTATAGGACCGAATA 247

    241 ysValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeu 257
          |||||||::   |||      |||:::      |||      ::::::
        248 CGGTCTTAACGGGTCAAAGAACCCCAACCCCAGGACCATCCTCTGATATA 297

    258 ProValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSe 274
                          :::   :::   |||   |||:::      ||
        298 ..................ACTTCTAAATTAGACCCCACTGAG......TC 323

    274 rAsnSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProG 291
        ||||||||||||
        324 TAACAGCACGACT..................................... 336

    291 lyValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAla 307
                  |||||||||   :::|||||||||||||||||||||||||
        337 ..........AAAACGGGGACAAAACTTTTTAGTCTCATCCAGGGAGCT 375

    308 PheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLe 324
        |||||||||:::|||||||||   |||:::|||||||||||||||||||||
        376 TTTCAAGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCT 425

    324 uCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluArgL 341
        |||||||:::|||||||||||||||||:::   |||||||||:::::|
        426 TTGCTTAACTTCGGGCCCACCTTACTATAAGAAAATGGCTAAAAGAGAAA 475

    341 ysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySerArg 357
        ||||||||||||||||:::||||||:::|||||||||||||||||||:::
        476 AATTCAATGTGACAAAAAAACATAGAGACCAATGTACATGGGGATCCCAA 525

    358 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLy 374
```

```
     |||||||||||||||||||||||||||||||||   |||||||||   ||
 526 AATAAGCTTACCCTTACTGAGGTTTCTGGAAAGGACACCTGCATAAAAAA 575

 374 sAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluG 391
     |:::|||||||||||||||||||||||      |||   :::::::::|
 576 GGTTCCCCCATCCCACCAACACCTTTACAACCACACTGAAGCCTTTAATC 625

 391 lnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAla 407
     ||:::||||||:::|||||||||||||||||||||:::|||||||||||||
 626 AAACCTCTGAGAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCA 675

 408 CysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSe 424
     |||||||||||||||||||||||||||||||||   ||||||||||||::
 676 TGTAATACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAAC 725

 424 rLysAspPheCysValMetValGlnIleValProArgValTyrTyrHisP 441
     :|||||||||   :::|||||||||||||||||||||||||||||||:::|
 726 TAAAGACTTTTACATTATGGTCCAAATTGTTCCCCGAGTATATTACTATC 775

 441 roGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgProLys 457
     ||:::::::::::::|||||||||||||||||||||   :::|||   |||
 776 CCAAGAAAACAATTCTCGATGAATATGATTACAGGAACCATCGACAAAAG 825

 458 ArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAl 474
     :::::::|||:::|||||||||||||||||||||||||||||||||:::
 826 AAAAAACCCATATCCCTGACACTCGCAGTAATGCTCGGACTCGGAGTGAT 875

 474 aValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnL 491
      :::||||||   ||||||||||||||||||:::||||||||||||||||
 876 AACAGGTGTGAGAACAGGAACTGCAGCTTTAGTTACAGGACCTCAGCAGC 925

 491 euGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArg 507
     |||||   ||||||:::::::||||||      :::|||   :::|||:::
 926 TAGAAACAGGACTTAGTAACCTACATCAAATTGTAACAGGAAATCTCCAA 975

 508 AlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSe 524
     |||||||||:::|||||||||||||||||||:::|||||||||||||||||
 976 GCCCTAAAAAAATCTGTCAGTAACCTGGAAAAATCCCTAACCTCCTTATC 1025

 524 rGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgG 541
     ||||||||||||||||||:::::|||||||||||||||||||||||::::
1026 TGAAGTAGTTCTACAGAATAAAAAAGGGTTAGATTTATTATTTCTAAAAA 1075

 541 luGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAsp 557
     ::      ||||||:::|||||||||||:::|||||||||||||||||||
1076 AAAGAAGATTATGTGTAGCCTTAAAGGAGAAATGCTGTTTTTATGTAGAT 1125

 558 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGl 574
     ||||||||||||||||||||||||||||||||||||:::|||:::|||||
1126 CATTCAGGGGCCATCAGAGACTCCATGAACAAGCTTAAAAAAAGGTTGGA 1175

 574 uArgArgArgArgGluArgGluAla.AspGlnGlyTrpPheGluGlyTrp 590
     |:::|||||||||||||:::|||:::    |||   ||||||:::   |||
1176 GAAACGTCGAAGGGAAAAGGAAACTTACTCAAAGATGGTTTAAAAGATGG 1225

 591 PheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyPr 607
     ||||||||||||||||||:::::|||||||||||||:::||||||||||||
1226 TTCAACAGGTCTCCTTGGTTGGCTACCCTACTTTCTACTTTAACAGGACC 1275

 607 oLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnA 624
     ||||:::|||||||||||||||||||||||||||||||||||:::|||||:
1276 CTTAATAGTCCTCCTCCTGTTACTCACAGTTGGGCCATGTATTATTAACA 1325

 624 rgPheValAlaPheValArgGluArgValSerAlaValGlnIleMetVal 640
     ::::::::|||||:::||:::|||:::|||||||||||||||||||||||
1326 AGTTAATTGCCTTCATTAGAAAACGAATAAGTGCAGTCCAGATCATGGTA 1375

 641 LeuArgGlnGlnTyrGln 646
     |||||||||||||||||
1376 CTTAGACAACAGTACCAA 1393

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-7
```

```
seq_documentation_block:
; Sequence 7, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 6312
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4058)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4059)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4060)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4061)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4246)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4247)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4248)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4249)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-7


alignment_scores:
            Quality: 1508.50              Length:    746
              Ratio:  3.060                Gaps:     24
 Percent Similarity:  66.086   Percent Identity: 44.102

alignment_block:
 US-09-171-553B-6 x US-09-011-745-7   ..

 Align seg 1/1  to: US-09-011-745-7  from: 1  to: 6312

       2 HisPro.................ThrLeuSerArgArgHisLeuProTh 12
         ||||||                 :::|||   |||||||||   |||::
    1428 CACCCGTTCCAGGTCGGGGATCAAGTGCTTGTCAGACGCCATCGACCCAG 1477
```


12 r.....ArgGlyGlyGluProLysArgLeuArgIleP 23
1478 CAGCCTTGAGCCTCGGTGGAAAGGCCCATACCTGGTGTTGCTGACTACCC 1527
23 roLeuSerPhe.....AlaSerIleAlaTrpPhe.LeuThrLeuTh 36
1528 CGACCGCGGTAAAAGTCGATGGTATTGCTGCCTGGGTCCATGCTTCTCAC 1577
36 r.....IleThrProGlnAlaSerSerL 44
1578 CTCAAACCTGCACCACCTTCGGCACCAGATGAGTCCT...GGGAGCTGGA 1624
44 ysArgLeuIle..... 47
1625 AAAGACTGATCATCCTCTTAAGCTGCGTATTCGGCGGCGGCGGGACGAGT 1674
48 ...AspSerSerAsnProHisArgProLeuSerLeuThrTrpLeuIleIl 63
1675 CTGCAAAATAAGAACCCCCACCAGCCCATGACCCTCACTTGGCAGGTACT 1724
63 eAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProA 80
1725 GTCC...CAAACTGGAGACGTTGTCTGGGATACAAAGGCAGTCCAGCCCC 1771
80 rgGlyThrTrpTrpProGluLeu.....HisPheCysLeuArgLeu... 93
1772 CTTGGACTTGGTGGCCCACACTTAAACCTGATGTATGTGCCTTGGCGGCT 1821
94 .....IleAsnProAlaValLy 99
1822 AGTCTTGAGTCCTGGGATATCCCGGGAACCGATGTCTCGTCCTCTAAACG 1871
99 sSerThrProProAsn..... 104
1872 AGTCAGACCTCCGGACTCAGACTATACTGCCGCTTATAAGCAAATCACCT 1921
105 .....LeuValArgSer 108
1922 GGGGAGCCATAGGGTGCAGCTACCCTCGGGCTAGGACTAGAATGGCAAGC 1971
109 TyrGlyPheTyrCysCysPro.....GlyThrGluLysGluLy 121
1972 TCTACCTTCTACGTATGTCCCCGGGATGGCCGGACCCTTTCAGAAGCTAG 2021
121 sTyrCysGlyGlySerGlyGluSerPheCysArgArgTrpSerCysValT 138
2022 AAGGTGCGGGGGGCTAGAATCCCTATACTGTAAAGAATGGGATTGTGAGA 2071
138 hrSerAsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgVal 154
2072 CCACGGGGACCGGTTATTGG...CTATCTAAATCCTCAAAAGACCTCATA 2118
155 LysPheSerPheValAsnSerGlyProGlyLysTyrLysMetMetLysLe 171
2119 ACTGTAAAATGGGACCAAAATAGCGAATGGACTCAAAAATTTCAACAGTG 2168
171 uTyrLysAspLysSerCysSerProSerAspLeuAspTyrLeuLysIleS 188
2169 TCACCAGACCGGCTGGTGTAACCCC.....CTTAAAATAG 2203
188 erPheThrGluArgLysThrGlyLysTyrSerLysValAspLysTrpTyr 204
2204 ATTTCACAGACAAA.....GGAAAATTATCCAAG.....GACTGGATA 2241
205 GluLeuGlyAsnSerPheLeuLeu.....TyrGlyGlyGlyAlaGl 218
2242 ...ACGGGAAAAACCTGGGGATTAAGATTCTATGTGTCTGGACATCCAGG 2288
218 ySerThrLeuThrIleArgLeuArgIleGluThrGlyThrGluProProV 235
2289 CGTACAGTTCACCATTCGCTTAAAAATCACC.....AACATGCCAGCTG 2332
235 alAlaMetGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeu 251
2333 TGGCAGTAGGTCCTGACCTCGTCCTTGTGGAACAAGGACCTCCTAGAACG 2382
252 Glu.....ProProHisAsnLeuProValProGl 261
2383 TCCCTCGCTCTCCCACCTCCTCTTCCCCCAAGGGAAGCGCCACCGCCATC 2432
261 nLeuThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrT 278
2433 TCTC.....CCCGACTCTAACTCCACAG 2455
278 hrGlyLeuIleProThrAsnThrProArg..... 287
2456 CCCTGGCGACTAGTGCACAAACTCCCACGGTGAGAAAAACAATTGTTACC 2505
288 ...AsnSerProGlyValProValLysThrGlyGlnArgLeuPheSerLe 303
2506 CTAAACACTCCG.....CCTCCCACCACAGGCGACAGACTTTTTGATCT 2549
303 uIleGlnGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrS 320
2550 TGTGCAGGGGGCCTTCCTAACCTTAAATGCTACCAACCCAGGGGCCACTG 2599
320 erSerCysTrpLeuCysLeuSerSerGlyProProTyrTyrGluGlyMet 336
2600 AGTCTTGCTGGCTTTGTTTGGCCATGGGCCCCCCTTATTATGAAGCAATA 2649
337 AlaLysGluArgLysPheAsnValThrLysGluHisArgAsnGlnCysTh 353
2650 GCCTCATCAGGAGAGGTCGCCTACTCCACCGAC...CTTGACCGGTGCCG 2696
353 rTrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyT 370
2697 CTGGGGGACCCAAGGAAAGCTCACCCTCACTGAGGTCTCAGGACACGGGT 2746
370 hrCysIleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThr 386
2747 TGTGCATAGGAAAGGTGCCCTTTACCCATCAGCATCTCTGCAATCAGACC 2796
387 ValValTyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAs 403
2797 CTATCCATCAATTCCTCCGGAGACCATCAGTATCTGCTCCCCTCCAACCA 2846
403 nArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerV 420
2847 TAGCTGGTGGGCTTGCAGCACTGGCCTCACCCCTTGCCTCTCCACCTCAG 2896
420 alPheAsnGlnSerLysAspPheCysValMetValGlnIleValProArg 436
2897 TTTTTAATCAGACTAGAGATTTCTGTATCCAGGTCCAGCTGATTCCTCGC 2946
437 ValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyrArgTy 453
2947 ATCTATTACTATCCTGAAGAAGTTTTGTTACAGGCCTATGACAATTCTCA 2996
453 rAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuG 470
2997 CCCCAGGACTAAAAGAGAGGCTGTCTCACTTACCCTAGCTGTTTTACTGG 3046
470 lyLeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThr 486
3047 GGTTGGGAATCACGGCGGGAATAGGTACTGGTTCAACTGCCTTAATTAAA 3096
487 GlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetTh 503
3097 GGACCTATAGACCTCCAGCAAGGCCTGACAAGCCTCCAGATCGCCATAGA 3146
503 rGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerL 520
3147 TGCTGACCTCCGGGCCCTCCAAGACTCAGTCAGCAAGTTAGAGGACTCAC 3196
520 euThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 536

```
                                                
```



seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-075-272-1

seq_documentation_block:
; Sequence 1, Application US/09075272
; Patent No. 6136598
; GENERAL INFORMATION:
; APPLICANT: MILLER, A. DUSTY
; APPLICANT: WOLGAMOT, GREG
; APPLICANT: BONHAM, LYNN
; TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL
; TITLE OF INVENTION: PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/075,272
; FILING DATE: 08-MAY-1998

```
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/046,140
;      FILING DATE:  09-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poor, Brian W.
;      REGISTRATION NUMBER:  32,928
;      REFERENCE/DOCKET NUMBER:  14538A-003710
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-075-272-1


alignment_scores:
           Quality: 1431.00              Length:    727
             Ratio:  3.019                Gaps:     25
 Percent Similarity:  65.199   Percent Identity: 43.466

alignment_block:
 US-09-171-553B-6 x US-09-075-272-1   ..

 Align seg 1/1  to: US-09-075-272-1  from: 1  to: 8655

      10 LeuProThrArgGlyGlyGluProLysArgLeuArgIlePro..... 23
         |||   |||      |||   |||   :::::::::   |||
    5724 CTTCACACGTCAAGAGGGCCCCCAGTCAAGATGAAGAAACCCACGAAGAC 5773

      24 ............................LeuSerPheAlaSerIleA 30
                                     ||||||||||||   :::|
    5774 AACTGGGCTGTGGAAGCCACTGATAACCCTCTTAAGCTTCGCTTGCGTCG 5823

      30 laTrpPheLeuThrLeuThrIleThrProGlnAlaSerSerLysArgLeu 46
         ||         ::::::|||:::
    5824 CAGGAGCCCCCTCCATCACCCTGGAC..................... 5849

      47 IleAspSerSerAsnProHisArgProLeuSerLeuThrTrpLeuIleIl 63
         :::   :::   |||||||||   |||::::::   :::|||   :::::
    5850 CTAGGGAACCACAACCCTCATGCCCCAGTTCAACAGTCTTGGGAAGTGCT 5899

      63 eAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProA 80
         ::::   :::   |||   :::|||      :::   :::|||   |||
    5900 TAAT...GAAAAGGGAGACGTTGTATGGGTAGCCACTGCAGTCCATCCCC 5946

      80 rgGlyThrTrpTrpProGluLeu......HisPheCys........... 90
            ||||||||||||:::|||         :::|||
    5947 CTTGGACTTGGTGGCCTGATCTCACACCTGACATCTGTAAATTAGCAGCA 5996

      91 ............................LeuArgLeuIleAsn...Pr 96
                                        |||   |||   ||
    5997 GGATCTCCCAATTGGGACCTTCCCGATCATACTGACTTGAATAACCCACC 6046

      96 oAlaValLysSerThrProProAsnLeuVal.................. 106
         |:::   ::::::      ||||||   |||
    6047 CTCTGAACAAAAGTGTGTCCCAAACGGGGTAGGAAGCACTACTGGATGTT 6096

     107 ........................ArgSerTyrGlyPheTyrCysCys 114
                                  |||:::     ||||||   |||
    6097 CGGGGCAGTTCTATCGAGCTAATCTTAGGGCTGCACAGTTTTATGTTTGC 6146

     115 ProGlyThr............GluLysGluLysTyrCysGlyGlySerGl 127
         ||||||               :::   ::::::   |||   |||:::::
    6147 CCTGGTCAGGGTCAGAAAGGAAAGCTGCAACAAGAATGTAGAGGGGCATC 6196

     127 yGluSerPheCysArgArgTrpSerCysValThrSerAsnAspGlyAspT 144
         ::::   ||||||   :::|||:::|||   |||:::::::::::   |

    6197 AGACTACTTTTGCGGTAAATGGACATGTGAAACAACAGGGGAAGCCTACT 6246

     144 rpLys............TrpProIle....................... 148
         |||||           |||   :::
    6247 GGAAGCCTTCCGCTGATTGGGACCTGATCACGGTAAAACGTGGTAGTGGT 6296

     149 ............SerLeuGlnAspArgValLysPheSerPheValAsnSe 161
                     :::      :::|||      ::::::::::::::||
    6297 TATGATAAGCCAAACCAAGGAGAAAGAAACCCATACAAATACCTAGATTC 6346

     161 rGly.......................ProGlyLysTyrLysMetMetL 170
         ||||                      ||||||
    6347 TGGGTGTGCTCTTAAAAATTACAGCCCCCCAGGACCA............. 6383

     170 ysLeuTyrLysAspLysSerCysSerProSerAspLeuAspTyrLeuLys 186
                |||   |||   |||:::|||               |||
    6384 .....TGCAAAGGTAAATACTGCAACCCC..............CTACTC 6413

     187 IleSerPheThrGluArgLysThrGlyLysTyrSerLysVal........ 200
         |||:::||||||||:::      ||||||   :::::::::
    6414 ATAAAATTCACTGAGAAA......GGGAAACAAGCTCGCCTGAGTTGGCT 6457

     201 .......AspLysTrpTyrGluLeuGlyAsnSerPheLeuLeuTyrGlyG 215
              :::::::|||         |||              :::
    6458 TAAAGGAAATAGGTGG.........GGTTGGCGAGTATATATTCCAATAA 6498

     215 lyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 231
              |||      :::||||||||||||   ::::::
    6499 GAGACCCTGGGTTTATCTTTACGATTAGACTGACAGTAAGA......GAC 6542

     232 GluProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyPr 248
              :::::::::||||||:::|||||||||:::||||||:::||
    6543 CTGGCAGTAACATCCATAGGACCCAACAAGGTCCTTACGGAACAGGCCCC 6592

     248 oProAlaLeuGluProProHisAsnLeuPro...ValProGlnLeuThrS 264
         ||||:::         |||   :::::::|||   ||||||         :
    6593 CCCAGTTGCACCGGCTCCCCCGAGAGTCCCAGCCGTGCCAGCTCCACCAA 6642

     264 erLeuArgProAspIleThrGlnProProSerAsnSerThrThrGlyLeu 280
         ::   ||||||      |||      ||||||   :::||||||
    6643 CTTCACGGCCC...TACACAGTAGGACCCTCATTAGAGACAACCTTAGCC 6689

     281 IleProThrAsnThrProArgAsnSerProGlyValProValLysThrGl 297
            |||                                 :::   |||
    6690 TCCCCACCA.........................CTCCTAGATACAGA 6712

     297 yGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSerT 314
          :::||||||   ||||||:::|||||||||||||   ::::::|||   |
    6713 AAACCGTCTGGTCAGTCTAGTTCAGGGAGCCTTTTTAGTTTTAAATAGGA 6762

     314 hrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyPro 330
         ||:::|||:::   |||:::|||||||||||||||   :::|||:::|||
    6763 CTAATCCTAATATGACTCAATCATGTTGGTTATGCTATGCCTCTAACCCC 6812

     331 ProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGl 347
         ||||||||||||||:::|||:::   |||   :::|||:::|||:::::
    6813 CCTTATTATGAAGGAATTGCTCAGACTAGAACTTACAATATTACTTCAGA 6862

     347 uHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThrG 364
         :|||   :::||||||   ||||||:::::::::|||||||||||||||
    6863 TCAT...TCTCAATGTCTTTGGGGAGAGAACAGGAAGTTGACTCTGACAG 6909

     364 luValSerGlyLysGlyThrCysIleGlyLysAlaProProSerHisGln 380
           |||||||||:::|||   |||:::|||:::::|||   :::
    6910 CAGTTTCAGGAAATGGGCTTTGTTTAGGTCAGGTGCCCCAGGATAAATGG 6959

     381 HisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsnGlnTy 397
         ||||||||   :::|||         :::      :::   :::|||||
    6960 CACCTCTGTAACCAGACTCAAAATATCCGACCTAACAAAGGTGGTCAGTA 7009

     397 rLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLeuThrP 414
         ||||||||||      :::      |||||||||||||||||||||||||
    7010 TCTAGTGCCTCCCATAGACACAGTATGGGCTTGCAATACAGGTCTCACTC 7059
```

```
     414 roCysValSerThrSerValPheAsnGlnSerLysAspPheCysValMet 430
         |||||:::|||   |||||||||||:::||||||||||||||:::::
    7060 CTTGTATATCTATGTCTGTTTTCAATAGCTCCAAAGATTTCTGTATTTTA 7109

     431 ValGlnIleValProArgValTyrTyrHisProGluGluValValLeuAs 447
         ||||||::::::||||||:::   ||||||   ::::::      |||||
    7110 GTTCAGCTTATTCCTAGACTCCTGTATCATGATGATAGCTCATTTCTAGA 7159

     447 pGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSerLeuT 464
         |:::::::::::|||      |||   ||||||||||||:::::::||||
    7160 CAAATTTGAACATCGGGTC...CGCTGGAAAAGAGAACCCATTACTTTAA 7206

     464 hrLeuAlaValMetLeuGlyLeuGly...ThrAlaValGlyValGlyThr 479
         ||||||||||:::||||||||||||   :::|||:::||||||||||||
    7207 CGTTAGCAGTGCTTTTAGGATTGGGAGTGGCCGCAGCTGGGGTAGGTACT 7256

     480 GlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGl 496
         ||||||||||||||||      |||:::   :::|||
    7257 GGAACCGCTGCCTTAATCCAGACCCCCCGATACTTTGAG.......... 7295

     496 yGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluServV 513
          ||||||:::::|||||      ||||||||||||:::::::::|||:
    7296 .GAATTACGTACAGCTATGGATACTGATCTCAGAGCTATAGAACACTCTA 7344

     513 alSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGln 529
         ::::::::||||||||||||||||||||||||||||||||||||||||
    7345 TAACCAAACTAGAAGAATCTTTAACTTCTCTGTCTGAGGTAGTACTGCAA 7394

     530 AsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAl 546
         ||||||||||||||||||||||||||||:::||||||||||||||||||
    7395 AATAGGAGAGGATTAGATTTGTTATTTCTTAAAGAAGGAGGACTTTGTGC 7444

     546 aAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyAlaIleA 563
         ||||||||||||||||||||||||||||||||||||||||||||:::|||:
    7445 TGCTCTCAAAGAGGAATGTTGTTTCTATGTCGACCACTCGGGAGTGATCA 7494

     563 rgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArgArgGlu 579
         ::|||||||||   |||||||||::::::|||:::   |||:::||||||
    7495 AAGATTCTATGGCCAAACTTAGAGAACGCCTAGATATACGTCAAAGAGAG 7544

     580 ArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTr 596
         ||||||:::   ||||||||||||||||:::|||||||||:::||||||||
    7545 AGAGAGAGCAAACAAGGATGGTTTGAAAGCTGGTTTAATAAGTCCCCCTG 7594

     596 pMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuL 613
         |:::||||||||||||||||:::::::::||||||||||:::::|||||||
    7595 GCTCACCACTCTCCTTTCCACTATAGCGGGACCCTTGATTATACTTCTGC 7644

     613 euLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheVal 629
         ||||||||||   |||||||||::::::|||:::::::||||||||:::
    7645 TTTTGCTTACTTTTGGCCCCTGCATTCTTAATAAGTTAGTAGCCTTTATT 7694

     630 ArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGl 646
         |||||||||::::::||||||||:::||||||||:::||||||||||||
    7695 AGAGAAAGAATAAATGCAGTACAGGTTATGGTATTAAAACAACAATATCA 7744

     646 nGlyLeuLeuSerGlnGlyGluThrAspLeu 656
         |   ::::::::::::::|||:::::|||
    7745 G...GTCTTCCAGGAGGCCGAAAACTCGCTC 7772

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-8

seq_documentation_block:
; Sequence 8, Application US/08110300A
; Patent No. 5643756
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,300A
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-110-300A-8


alignment_scores:
            Quality: 1299.00              Length:    709
              Ratio:  2.855                Gaps:     26
 Percent Similarity:  64.175   Percent Identity: 42.595

alignment_block:
 US-09-171-553B-6 x US-08-110-300A-8   ..

 Align seg 1/1  to: US-08-110-300A-8  from: 1  to: 8323

       4 ThrLeuSerArgArgHisLeuProThrArgGlyGlyGluProLysArgLe 20
         ||||||||||:::         |||            :::|||:::   ||
    5787 ACGCTCTCAAAATCC......CCTAAAGATAAGATTGACCCGCGGGACCT 5830

      20 uArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrI 37
         |   |||||||||   :::                |||||:::|||
    5831 CCTAATCCCCTTAATTCTC..............TTCCTGTCTCTC.... 5861

      37 leThrProGlnAlaSerSerLysArgLeuIleAspSerSerAsnProHis 53
                        :::::::::           :::|||:::||||||
    5862 ..........AAAGGGGCCAGATCCGCAGCACCCGGCTCCAGCCCTCAC 5900

      54 ArgProLeuSerLeuThrTrpLeuIleIleAspProAspThrGlyValTh 70
         :::      :::::||||||   :::   :::   |||          ||
    5901 CAGGTCTACAACATTACCTGGGAAGTGACCAATGGGGATCGG...GAGAC 5947

      70 rValAsnSerThrArgGlyValAlaProArgGlyThrTrpTrpProGluL 87
         ||||   :::      |||      |||      ||||||||||||   |
    5948 AGTATGGGCAATATCAGGCAACCACCCTCTGTGGACTTGGTGGCCAGTCC 5997

      87 eu.....................................HisPhe... 89
         ||                                    |||:::
    5998 TCACCCCAGATTTGTGTATGTTAGCTCTCAGTGGGCCGCCCCACTGGGGG 6047

      89 ..................................................... 89

    6048 CTAGAGTATCAGGCCCCCTATTCCTCGCCCCCGGGGCCCCCTTGTTGCTC 6097

      90 .......................CysLeuArgLeuIleAsnProAlaV 98
                                |||   |||      |||      :
```

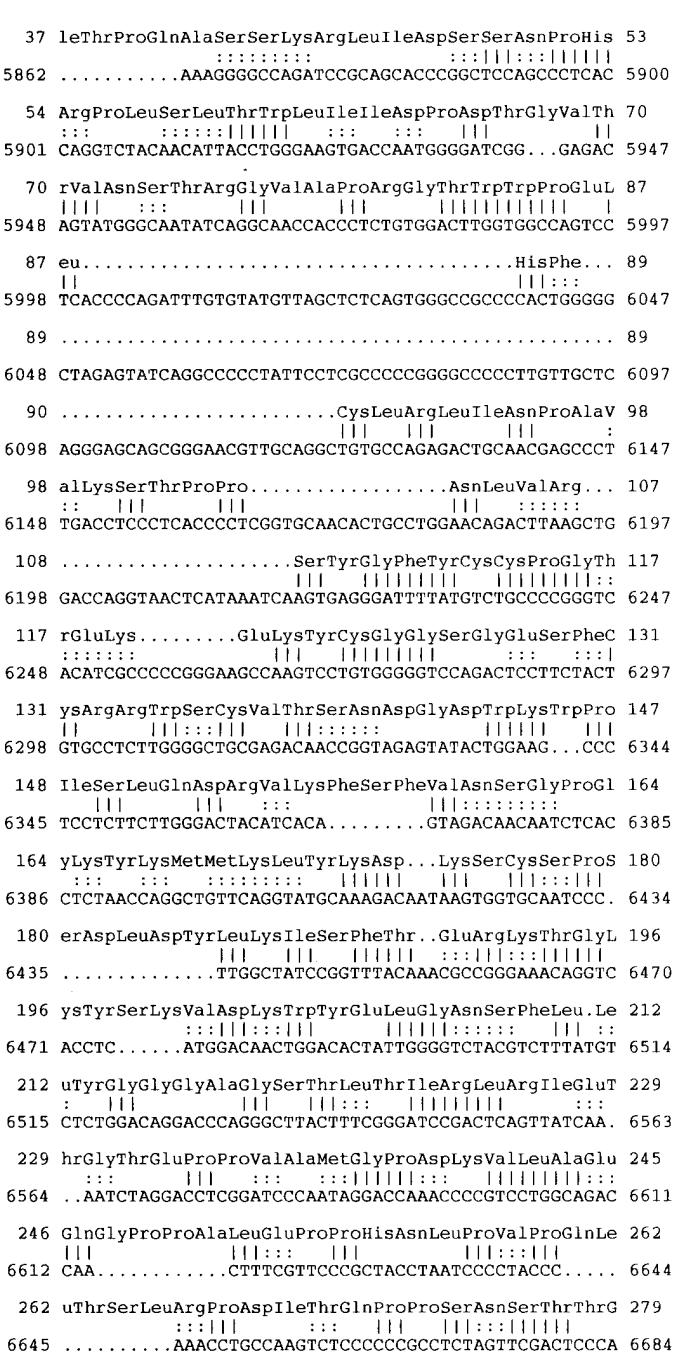
```
6098 AGGGAGCAGCGGGAACGTTGCAGGCTGTGCCAGAGACTGCAACGAGCCCT 6147

  98 alLysSerThrProPro.................AsnLeuValArg... 107
     ::   |||      |||                  |||   ::::::
6148 TGACCTCCCTCACCCCTCGGTGCAACACTGCCTGGAACAGACTTAAGCTG 6197

 108 .....................SerTyrGlyPheTyrCysCysProGlyTh 117
                          |||   |||||||||   |||||||||::
6198 GACCAGGTAACTCATAAATCAAGTGAGGGATTTTATGTCTGCCCCGGGTC 6247

 117 rGluLys.........GluLysTyrCysGlyGlySerGlyGluSerPheC 131
     :::::::             |||   |||||||||      :::   :::|
6248 ACATCGCCCCCGGGAAGCCAAGTCCTGTGGGGGTCCAGACTCCTTCTACT 6297

 131 ysArgArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpPro 147
     ||      |||:::|||   |||::::::          ||||||   |||
6298 GTGCCTCTTGGGGCTGCGAGACAACCGGTAGAGTATACTGGAAG...CCC 6344

 148 IleSerLeuGlnAspArgValLysPheSerPheValAsnSerGlyProGl 164
        |||      |||   :::             |||::::::::::
6345 TCCTCTTCTTGGGACTACATCACA.........GTAGACAACAATCTCAC 6385

 164 yLysTyrLysMetMetLysLeuTyrLysAsp...LysSerCysSerProS 180
      :::   :::   :::::::::  ||||||   |||   |||:::|||
6386 CTCTAACCAGGCTGTTCAGGTATGCAAAGACAATAAGTGGTGCAATCCC. 6434

 180 erAspLeuAspTyrLeuLysIleSerPheThr..GluArgLysThrGlyL 196
                  |||   |||   ||||||  :::|||:::||||||
6435 ..............TTGGCTATCCGGTTTACAAACGCCGGGAAACAGGTC 6470

 196 ysTyrSerLysValAspLysTrpTyrGluLeuGlyAsnSerPheLeu.Le 212
               :::|||:::|||      ||||||::::::   ||| ::
6471 ACCTC......ATGGACAACTGGACACTATTGGGGTCTACGTCTTTATGT 6514

 212 uTyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGluT 229
     :  |||         |||   |||:::   |||||||||      :::
6515 CTCTGGACAGGACCCAGGGCTTACTTTCGGGATCCGACTCAGTTATCAA. 6563

 229 hrGlyThrGluProProValAlaMetGlyProAspLysValLeuAlaGlu 245
       :::      |||   :::   :::|||||||:::   ||||||||:::
6564 ..AATCTAGGACCTCGGATCCCAATAGGACCAAACCCCGTCCTGGCAGAC 6611

 246 GlnGlyProProAlaLeuGluProProHisAsnLeuProValProGlnLe 262
     |||            |||:::   |||         |||:::|||
6612 CAA............CTTTCGTTCCCGCTACCTAATCCCCTACCC..... 6644

 262 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrG 279
              :::|||      :::   |||   |||:::||||||
6645 ..........AAACCTGCCAAGTCTCCCCCCGCCTCTAGTTCGACTCCCA 6684

 279 lyLeuIle...ProThrAsnThrProArgAsnSerProGlyValProVal 294
       ||||||   |||:::   ||||||  :::   |||      |||:::
6685 CATTGATTTCCCCGTCCCCCACTCCCACTCAGCCCCCG......CCAGCA 6728

 295 LysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIl 311
        ||||||:::||||||:::::::|||:::|||||||||:::||||||::
6729 GGAACGGGAGACAGATTACTAAATCTAGTACAGGGAGCTTACCAGGCACT 6778

 311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerS 328
     :|||   |||:::||||||   |||::::::|||||||||||||||   |
6779 CAACCTTACCAACCCTGATAAAACTCAAGAGTGCTGGTTATGCCTAGTGT 6828

 328 erGlyProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnVal 344
     ||||||||||||||||||||||:::|||           ::::::
6829 CTGGACCCCCCTATTACGAGGGGGTTGCCGTCCTAGGTACTTATTCCAAC 6878

 345 ThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuTh 361
           :::         :::|||:::   :::|||::::::|||||||
6879 CATACCTCTGCCCCAGCTAACTGCTCCGTGGCCTCCCAACACAAGCTGAC 6928

 361 rLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProProS 378
     ||||:::||||||:::|||:::|||   |||||||||   :::|||   :
6929 CCTGTCCGAAGTGACTGGACGGGGACTCTGCATAGGAACAGTCCCAAAAA 6978

 378 erHisGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGlu 394
     ::|||||||   ||||||   :::|||:::      :::   ::::::
6979 CTCACCAGGCCCTGTGCAACACTACCCTT......AAGGCAGGCAAAGGG 7022

 395 AsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGl 411
     :::   |||||||||         :::      ||||||||||||||||||
7023 TCTTACTATCTAGTTGCCCCCACAGGAACTATGTGGGCATGTAACACTGG 7072

 411 yLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheC 428
     ||||||||||||:::|||::::::|||:::|||::::::   |||:::|
7073 ACTCACTCCATGCCTATCTGCCACCGTGCTTAATCGCACCACTGACTATT 7122

 428 ysValMetValGlnIleValProArgValTyrTyrHisProGluGluVal 444
     ||||||:::|||::::::   |||||||||   |||||||||   :::
7123 GCGTTCTCGTGGAATTATGGCCCAGGGTCACCTACCATCCTCCCAGTTAC 7172

 445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProVa 461
     |||   ::::::::::::   :::   :::|||   ||||||||||||||
7173 GTCTATAGCCAGTTTGAA...AAATCCCATAGACATAAAAGAGAACCAGT 7219

 461 lSerLeuThrLeuAlaValMetLeuGly............LeuGlyThrA 474
     ||||||||||||||||:::::|||||                |||   |
7220 GTCCTTAACCTTGGCCTTATTATTAGGTGGGCTAACTATGGGTGGCATCG 7269

 474 laValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGln 490
     ||:::|||||||||||||||||||:::|||||:::   :::   ||||||
7270 CCGCGGGAGTAGGGACAGGAACTACCGCCCTGGTC...GCCACCCAGCAG 7316

 491 LeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuAr 507
     ::::::            :::||||||||||||:::   :::|||||||::
7317 TTTCAG............CAGCTCCATGCTGCCGTACAAGATGATCTCAA 7354

 507 gAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuS 524
     :   ::::::::::|||::::::|||||||||:::||||||||||||||||
7355 AGAAGTCGAAAAGTCAATTACTAACCTAGAAAAGTCTCTTACTTCGTTGT 7404

 524 erGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArg 540
     ||||||||||||||||||||||||||||||||||||||||||||||:::
7405 CTGAGGTTGTACTGCAGAATCGACGAGGCCTAGACCTGTTGTTCCTAAAA 7454

 541 GluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAs 557
     ||||||||||||||||||||||||||||||||||||||||||||:::||
7455 GAGGGAGGACTGTGTGCTGCCCTAAAAGAAGAATGTTGTTTCTATGCTGA 7504

 557 pHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuG 574
     ||||:::|||   :::||||||||||||   |||||||||::::::|||:
7505 CCATACAGGCCTAGTAAGAGATAGTATGGCCAAATTAAGAGAGAGACTCT 7554

 574 luArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrp 590
     ::::::|||::::::       |||:::::::||||||||||||||||||||
7555 CTCAGAGACAAAAACTATTTGAGTCGAGCCAAGGATGGTTCGAAGGATGG 7604

 591 PheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyPr 607
     ||||||||||||||||||:::|||||||:::|||::::::   |||||
7605 TTTAACAGATCCCCCTGGTTTACCACGTTGATATCCACCATCATGGGGCC 7654

 607 oLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnA 624
     ||||::::::|||||||||:::|||       |||||||||::::::|||||
7655 TCTCATTATACTCCTACTAATTCTGCTTTTTGGACCCTGCATTCTTAATC 7704

 624 rgPheValAlaPheValArgGluArgValSerAlaValGlnIleMetVal 640
     ||:::|||   ||||||::::::|||:::|||:::||||||   :::|||
7705 GATTAGTTCAATTTGTTAAAGACAGGATCTCAGTAGTCCAGGCTTTAGTC 7754

 641 LeuArgGlnGlnTyrGlnGlyLeu 648
     |||   |||||||||:::   |||
7755 CTGACTCAACAATACCACCAGCTA 7778
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-8

seq_documentation_block:
; Sequence 8, Application US/08886642

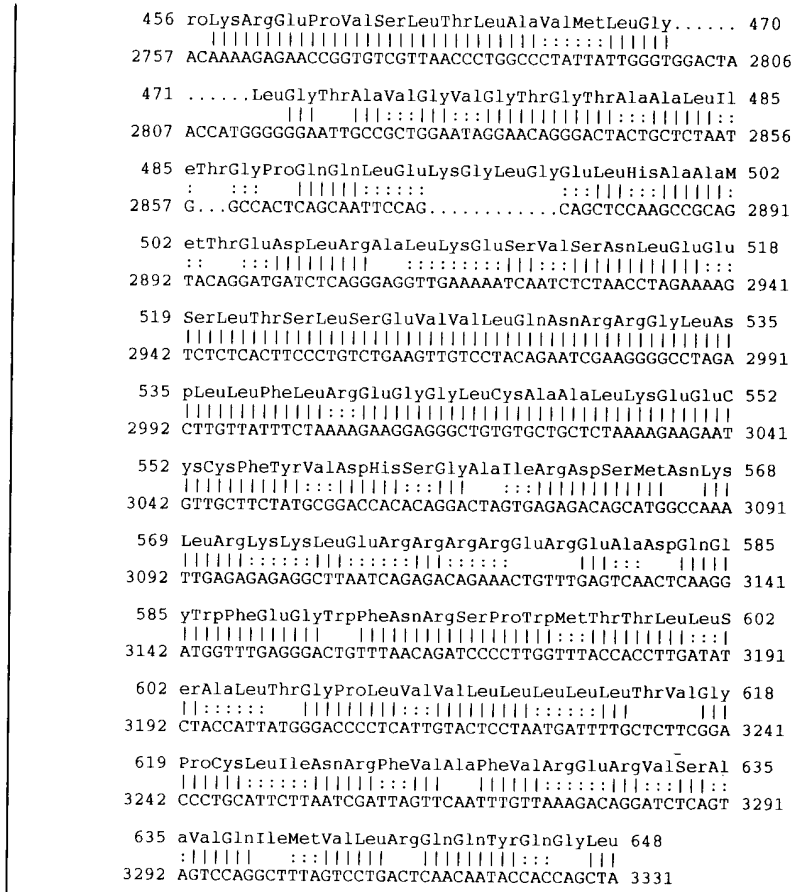
```
; Patent No. 5952474
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  45 Rockefeller Plaza, Suite 2800
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/886,642
;      FILING DATE:  01-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/110,300
;      FILING DATE:  20-AUG-1993
;      APPLICATION NUMBER:  07/938,100
;      FILING DATE:  28-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hone, William J.
;      REGISTRATION NUMBER:  26,739
;      REFERENCE/DOCKET NUMBER:  07763/010002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212/765-5070
;      TELEFAX:  212/258-2291
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-886-642-8


alignment_scores:
            Quality: 1299.00              Length:    709
              Ratio:  2.855                Gaps:     26
 Percent Similarity:  64.175   Percent Identity: 42.595

alignment_block:
 US-09-171-553B-6 x US-08-886-642-8   ..

 Align seg 1/1  to: US-08-886-642-8  from: 1  to: 8323

       4 ThrLeuSerArgArgHisLeuProThrArgGlyGlyGluProLysArgLe 20
         ||||||||||:::         |||           :::|||:::   ||
    5787 ACGCTCTCAAAATCC......CCTAAAGATAAGATTGACCCGCGGGACCT 5830

      20 uArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrI 37
         |   |||||||||   :::               ||||||:::|||
    5831 CCTAATCCCCTTAATTCTC..............TTCCTGTCTCTC.... 5861

      37 leThrProGlnAlaSerSerLysArgLeuIleAspSerSerAsnProHis 53
                    :::::::::           :::|||:::||||||
    5862 ...........AAAGGGGCCAGATCCGCAGCACCCGGCTCCAGCCCTCAC 5900

      54 ArgProLeuSerLeuThrTrpLeuIleIleAspProAspThrGlyValTh 70
         :::      ::::::||||||   :::   :::   |||         ||
    5901 CAGGTCTACAACATTACCTGGGAAGTGACCAATGGGGATCGG...GAGAC 5947

      70 rValAsnSerThrArgGlyValAlaProArgGlyThrTrpTrpProGluL 87
         ||||   :::      |||      |||      ||||||||||||   |
    5948 AGTATGGGCAATATCAGGCAACCACCCTCTGTGGACTTGGTGGCCAGTCC 5997

      87 eu.....HisPhe... 89
         ||          |||:::
    5998 TCACCCCAGATTTGTGTATGTTAGCTCTCAGTGGGCCGCCCCACTGGGGG 6047

      89 ..... 89

    6048 CTAGAGTATCAGGCCCCCTATTCCTCGCCCCCGGGGCCCCCTTGTTGCTC 6097

      90 .....CysLeuArgLeuIleAsnProAlaV 98
                |||   |||      |||      :
    6098 AGGGAGCAGCGGGAACGTTGCAGGCTGTGCCAGAGACTGCAACGAGCCCT 6147

      98 alLysSerThrProPro.....AsnLeuValArg... 107
         ::   |||      |||          |||   ::::::
    6148 TGACCTCCCTCACCCCTCGGTGCAACACTGCCTGGAACAGACTTAAGCTG 6197

     108 .....SerTyrGlyPheTyrCysCysProGlyTh 117
                    |||   |||||||||   |||||||||::
    6198 GACCAGGTAACTCATAAATCAAGTGAGGGATTTTATGTCTGCCCCGGGTC 6247

     117 rGluLys.....GluLysTyrCysGlyGlySerGlyGluSerPheC 131
         :::::::          |||   |||||||||      :::   :::|
    6248 ACATCGCCCCCGGGAAGCCAAGTCCTGTGGGGGTCCAGACTCCTTCTACT 6297

     131 ysArgArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpPro 147
         ||      |||:::|||   |||::::::          ||||||   |||
    6298 GTGCCTCTTGGGGCTGCGAGACAACCGGTAGAGTATACTGGAAG...CCC 6344

     148 IleSerLeuGlnAspArgValLysPheSerPheValAsnSerGlyProGl 164
            |||      |||   :::            |||:::::::::
    6345 TCCTCTTCTTGGGACTACATCACA.........GTAGACAACAATCTCAC 6385

     164 yLysTyrLysMetMetLysLeuTyrLysAsp...LysSerCysSerProS 180
          :::   :::   :::::::::  ||||||   |||   |||:::|||
    6386 CTCTAACCAGGCTGTTCAGGTATGCAAAGACAATAAGTGGTGCAATCCC. 6434

     180 erAspLeuAspTyrLeuLysIleSerPheThr..GluArgLysThrGlyL 196
                       |||   |||   ||||||  :::|||:::||||||
    6435 .............TTGGCTATCCGGTTTACAAACGCCGGGAAACAGGTC 6470

     196 ysTyrSerLysValAspLysTrpTyrGluLeuGlyAsnSerPheLeu.Le 212
                    :::|||:::|||      ||||||::::::   ||| ::
    6471 ACCTC......ATGGACAACTGGACACTATTGGGGTCTACGTCTTTATGT 6514

     212 uTyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGluT 229
         :  |||          |||   |||:::  |||||||||      :::
    6515 CTCTGGACAGGACCCAGGGCTTACTTTCGGGATCCGACTCAGTTATCAA. 6563

     229 hrGlyThrGluProProValAlaMetGlyProAspLysValLeuAlaGlu 245
           :::       |||   :::   :::|||||||:::   |||||||||:::
    6564 ..AATCTAGGACCTCGGATCCCAATAGGACCAAACCCCGTCCTGGCAGAC 6611

     246 GlnGlyProProAlaLeuGluProProHisAsnLeuProValProGlnLe 262
         |||            |||:::   |||         |||:::|||
    6612 CAA............CTTTCGTTCCCGCTACCTAATCCCCTACCC..... 6644

     262 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrG 279
                 :::|||      :::   |||   |||:::||||||
    6645 ..........AAACCTGCCAAGTCTCCCCCCGCCTCTAGTTCGACTCCCA 6684

     279 lyLeuIle...ProThrAsnThrProArgAsnSerProGlyValProVal 294
           ||||||   |||:::   ||||||   :::   |||       |||:::
    6685 CATTGATTTCCCCGTCCCCCACTCCCACTCAGCCCCCG......CCAGCA 6728

     295 LysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIl 311
            ||||||:::||||||:::::::|||:::||||||||||:::||||||::
    6729 GGAACGGGAGACAGATTACTAAATCTAGTACAGGGAGCTTACCAGGCACT 6778

     311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerS 328
         :|||   |||:::||||||   |||:::::::||||||||||||||   |
    6779 CAACCTTACCAACCCTGATAAAACTCAAGAGTGCTGGTTATGCCTAGTGT 6828

     328 erGlyProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnVal 344
```

```
                                                
```


37 leThrProGlnAlaSerSerLysArgLeuIleAspSerSerAsnProHis 53
5862 .....AAAGGGGCCAGATCCGCAGCACCCGGCTCCAGCCCTCAC 5900
54 ArgProLeuSerLeuThrTrpLeuIleIleAspProAspThrGlyValTh 70
5901 CAGGTCTACAACATTACCTGGGAAGTGACCAATGGGGATCGG...GAGAC 5947
70 rValAsnSerThrArgGlyValAlaProArgGlyThrTrpTrpProGluL 87
5948 AGTATGGGCAATATCAGGCAACCACCCTCTGTGGACTTGGTGGCCAGTCC 5997
87 eu.....HisPhe... 89
5998 TCACCCCAGATTTGTGTATGTTAGCTCTCAGTGGGCCGCCCCACTGGGGG 6047
89 89
6048 CTAGAGTATCAGGCCCCCTATTCCTCGCCCCCGGGGCCCCCTTGTTGCTC 6097
90 .....CysLeuArgLeuIleAsnProAlaV 98
6098 AGGGAGCAGCGGGAACGTTGCAGGCTGTGCCAGAGACTGCAACGAGCCCT 6147
98 alLysSerThrProPro.....AsnLeuValArg... 107
6148 TGACCTCCCTCACCCCTCGGTGCAACACTGCCTGGAACAGACTTAAGCTG 6197
108 .....SerTyrGlyPheTyrCysCysProGlyTh 117
6198 GACCAGGTAACTCATAAATCAAGTGAGGGATTTTATGTCTGCCCCGGGTC 6247
117 rGluLys.....GluLysTyrCysGlyGlySerGlyGluSerPheC 131
6248 ACATCGCCCCCGGGAAGCCAAGTCCTGTGGGGGTCCAGACTCCTTCTACT 6297
131 ysArgArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpPro 147
6298 GTGCCTCTTGGGGCTGCGAGACAACCGGTAGAGTATACTGGAAG...CCC 6344
148 IleSerLeuGlnAspArgValLysPheSerPheValAsnSerGlyProGl 164
6345 TCCTCTTCTTGGGACTACATCACA.....GTAGACAACAATCTCAC 6385
164 yLysTyrLysMetMetLysLeuTyrLysAsp...LysSerCysSerProS 180
6386 CTCTAACCAGGCTGTTCAGGTATGCAAAGACAATAAGTGGTGCAATCCC. 6434
180 erAspLeuAspTyrLeuLysIleSerPheThr..GluArgLysThrGlyL 196
6435 .....TTGGCTATCCGGTTTACAAACGCCGGGAAACAGGTC 6470
196 ysTyrSerLysValAspLysTrpTyrGluLeuGlyAsnSerPheLeu.Le 212
6471 ACCTC.....ATGGACAACTGGACACTATTGGGGTCTACGTCTTTATGT 6514
212 uTyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGluT 229
6515 CTCTGGACAGGACCCAGGGCTTACTTTCGGGATCCGACTCAGTTATCAA. 6563
229 hrGlyThrGluProProValAlaMetGlyProAspLysValLeuAlaGlu 245
6564 ..AATCTAGGACCTCGGATCCCAATAGGACCAAACCCCGTCCTGGCAGAC 6611
246 GlnGlyProProAlaLeuGluProProHisAsnLeuProValProGlnLe 262
6612 CAA.....CTTTCGTTCCCGCTACCTAATCCCCTACCC..... 6644
262 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrG 279
6645 .....AAACCTGCCAAGTCTCCCCCCGCCTCTAGTTCGACTCCCA 6684
279 lyLeuIle...ProThrAsnThrProArgAsnSerProGlyValProVal 294
6685 CATTGATTTCCCCGTCCCCCACTCCCACTCAGCCCCCG.....CCAGCA 6728
295 LysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIl 311
6729 GGAACGGGAGACAGATTACTAAATCTAGTACAGGGAGCTTACCAGGCACT 6778
311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerS 328
6779 CAACCTTACCAACCCTGATAAAACTCAAGAGTGCTGGTTATGCCTAGTGT 6828
328 erGlyProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnVal 344
6829 CTGGACCCCCCTATTACGAGGGGGTTGCCGTCCTAGGTACTTATTCCAAC 6878
345 ThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuTh 361
6879 CATACCTCTGCCCCAGCTAACTGCTCCGTGGCCTCCCAACACAAGCTGAC 6928
361 rLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProProS 378
6929 CCTGTCCGAAGTGACTGGACGGGGACTCTGCATAGGAACAGTCCCAAAAA 6978
378 erHisGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGlu 394
6979 CTCACCAGGCCCTGTGCAACACTACCCTT.....AAGGCAGGCAAAGGG 7022
395 AsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGl 411
7023 TCTTACTATCTAGTTGCCCCCACAGGAACTATGTGGGCATGTAACACTGG 7072
411 yLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheC 428
7073 ACTCACTCCATGCCTATCTGCCACCGTGCTTAATCGCACCACTGACTATT 7122
428 ysValMetValGlnIleValProArgValTyrTyrHisProGluGluVal 444
7123 GCGTTCTCGTGGAATTATGGCCCAGGGTCACCTACCATCCTCCCAGTTAC 7172
445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProVa 461
7173 GTCTATAGCCAGTTTGAA...AAATCCCATAGACATAAAAGAGAACCAGT 7219
461 lSerLeuThrLeuAlaValMetLeuGly.....LeuGlyThrA 474
7220 GTCCTTAACCTTGGCCTTATTATTAGGTGGGCTAACTATGGGTGGCATCG 7269
474 laValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGln 490
7270 CCGCGGGAGTAGGGACAGGAACTACCGCCCTGGTC...GCCACCCAGCAG 7316
491 LeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuAr 507
7317 TTTCAG.....CAGCTCCATGCTGCCGTACAAGATGATCTCAA 7354
507 gAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuS 524
7355 AGAAGTCGAAAAGTCAATTACTAACCTAGAAAAGTCTCTTACTTCGTTGT 7404
524 erGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArg 540
7405 CTGAGGTTGTACTGCAGAATCGACGAGGCCTAGACCTGTTGTTCCTAAAA 7454
541 GluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAs 557
7455 GAGGGAGGACTGTGTGCTGCCCTAAAAGAAGAATGTTGTTTCTATGCTGA 7504
557 pHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuG 574
7505 CCATACAGGCCTAGTAAGAGATAGTATGGCCAAATTAAGAGAGAGACTCT 7554
574 luArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrp 590

```
      :::::|||::::::      |||::::::|||||||||||||||||||||
   7555 CTCAGAGACAAAAACTATTTGAGTCGAGCCAAGGATGGTTCGAAGGATGG 7604

    591 PheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyPr 607
        ||||||||||||||||||:::|||||||||:::|||::::::   |||||
   7605 TTTAACAGATCCCCCTGGTTTACCACGTTGATATCCACCATCATGGGGCC 7654

    607 oLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnA 624
        ||||::::::|||||||||:::|||      |||||||||::::::||||
   7655 TCTCATTATACTCCTACTAATTCTGCTTTTTGGACCCTGCATTCTTAATC 7704

    624 rgPheValAlaPheValArgGluArgValSerAlaValGlnIleMetVal 640
        ||:::|||   ||||||:::::|||:::|||:::||||||   :::|||
   7705 GATTAGTTCAATTTGTTAAAGACAGGATCTCAGTAGTCCAGGCTTTAGTC 7754

    641 LeuArgGlnGlnTyrGlnGlyLeu 648
        |||   |||||||||:::   |||
   7755 CTGACTCAACAATACCACCAGCTA 7778

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-6

seq_documentation_block:
; Sequence 6, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 6061
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3807)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3808)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3809)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3810)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3995)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3996)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3997)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3998)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-6


alignment_scores:
            Quality: 1258.00              Length:    730
              Ratio:  2.759                Gaps:     26
 Percent Similarity:  62.466   Percent Identity: 40.822

alignment_block:
 US-09-171-553B-6 x US-09-011-745-6   ..

 Align seg 1/1  to: US-09-011-745-6  from: 1  to: 6061

       2 HisProThrLeuSerArgArgHis.................LeuProTh 12
         |||         ||||||||||||                    |||
    1265 CACCGCCCTCAAAGTAGACGGCATCGCAGCTTGGATACACGCCGCCCACG 1314

      12 rArgGlyGlyGluProLysArgLeuArgIle.........ProLeuSerP 26
          :::|||   :::|||:::         |||            ::::::
    1315 TGAAGGCTGCCGACCCCGGGGGTGGACCATCCTCTAGACTGACATGGCGC 1364

      26 heAlaSerIleAlaTrpPheLeuThrLeuThrIleThrProGlnAlaSer 42
           ::::::::::::     |||         ::::::|||::::::
    1365 GTTCAACGCTCTCAAAACCCCTTAAAAATAAGGTTAACCCGCGAGGCCCC 1414

      43 SerLysArgLeuIle..............................As 48
                 ||||||
    1415 CTAATCCCCTTAATTCTTCTGATGCTCAGAGGGGTCAGTACTGCTTCGCC 1464

      48 pSerSerAsnProHisArgProLeuSerLeuThrTrpLeuIleIleAspP 65
          :::|||:::|||||||:::      ::::::||||||   :::   :::
    1465 CGGCTCCAGTCCTCATCAAGTCTATAATATCACCTGGGAGGTAACCAATG 1514

      65 roAspThrGlyValThrValAsnSerThrArgGlyValAlaProArgGly 81
           |||         ||||||   :::|||   |||      |||
    1515 GAGATCGG...GAGACGGTATGGGCAACTTCTGGCAACCACCCTCTGTGG 1561

      82 ThrTrpTrpProGluLeu.............................. 87
         ||||||||||||:::|||
    1562 ACCTGGTGGCCTGACCTTACCCCAGATTTATGTATGTTAGCCCACCATGG 1611

      88 .......HisPheCysLeuArgLeuIleAsnProAlaValLysSerThrP 102
                ::::::   |||:::      :::|||         |||:::|
    1612 ACCATCTTATTGGGGGCTAGAATATCAATCCCCTTTT......TCTTCTC 1655

     102 roPro.......................................... 103
         |||||
    1656 CCCCGGGGCCCCCTTGTTGCTCAGGGGGCAGCAGCCCAGGCTGTTCCAGA 1705

     103 ............................................... 103

    1706 GACTGCGAAGAACCTTTAACCTCCCTCACCCCTCGGTGCAACACTGCCTG 1755

     104 .AsnLeuValArg.....................SerTyrGlyPheT 112
          |||   ::::::                        :::   |||||||
    1756 GAACAGACTCAAGCTAGACCAGACAACTCATAAATCAAATGAGGGATTTT 1805

     112 yrCysCysProGlyThrGluLys.........GluLysTyrCysGlyGly 125
         ||   |||||||||   ::::::         :::|||   |||||||||
    1806 ATGTTTGCCCCGGGCCCCACCGCCCCCGAGAATCCAAGTCATGTGGGGGT 1855

     126 SerGlyGluSerPheCysArgArgTrpSerCysValThrSerAsnAspGl 142
               :::   :::|||      |||:::|||   |||::::::   ::
    1856 CCAGACTCCTTCTACTGTGCCTATTGGGGCTGTGAGACAACCGGTAGAGC 1905

     142 yAspTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheV 159
         :   ||||||   |||   |||      |||          |||      |
    1906 TTACTGGAAG...CCCTCCTCATCATGGGAT.........TTCATCACAG 1943
```

159 alAsnSerGlyProGlyLysTyrLysMetMetLysLeuTyrLysAsp... 174
1944 TAAACAACAATCTCACCTCTGACCAGGCTGTCCAGGTATGCAAAGATAAT 1993
175 LysSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGl 191
1994 AAGTGGTGCAACCCC.....TTAGTTATTCGGTTTACAGA 2028
191 u......ArgLysThrGlyLysTyrSerLysValAspLysTrpTyrGluL 206
2029 CGCCGGGAGACGGGTTACTTCCTGGACCACAGGACATTACTGG..... 2071
206 euGlyAsnSerPheLeuLeuTyrGlyGlyGlyAlaGlySerThrLeuThr 222
2072 ..GGCTTACGTTTGTATGTCTCCGGACAAGATCCAGGGCTTACATTTGGG 2119
223 IleArgLeuArgIleGluThrGlyThrGluProProValAlaMetGlyPr 239
2120 ATCCGACTCAGATACCAA...AATCTAGGACCCCGCGTCCCAATAGGGCC 2166
239 oAspLysValLeuAlaGluGlnGlyProProAlaLeuGluProProHisA 256
2167 AAACCCCGTTCTGGCAGACCAACAGCCA......CTCTCCAAGCCCAAA. 2209
256 snLeuProValProGlnLeuThrSerLeuArgProAspIleThrGlnPro 272
2210 .....CCTGTTAAGTCG.....CCTTCAGTCACCAAACCA 2239
273 ProSerAsnSerThrThrGlyLeuIleProThrAsnThrProArgAsnSe 289
2240 CCCAGTGGG......ACTCCTCTCTCCCCTACCCAACTTCCA..... 2275
289 rProGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnG 306
2276 .....CCGGCGGGAACGGAAAATAGGCTGCTAAACTTAGTAGACG 2315
306 lyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCys 322
2316 GAGCCTACCAAGCCCTCAACCTCACCAGTCCTGACAAAACCCAAGAGTGC 2365
323 TrpLeuCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGl 339
2366 TGGTTGTGTCTAGTAGCGGGACCCCCCTACTACGAAGGGGTTGCCGTCCT 2415
339 uArgLysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlyS 356
2416 GGGTACCTACTCCAACCATACCTCTGCTCCAGCCAACTGCTCCGTGGCCT 2465
356 erArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIle 372
2466 CCCAACACAAGTTGACCCTGTCCGAAGTGACCGGACAGGGACTCTGCATA 2515
373 GlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTy 389
2516 GGAGCAGTTCCCAAAACACATCAGGCCCTATGTAATACCACCCAGACAAG 2565
389 rGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpT 406
2566 CAGTCGAGGGTCC......TATTATCTAGTTGCCCCTACAGGTACCATGT 2609
406 rpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsn 422
2610 GGGCTTGTAGTACCGGGCTTACTCCATGCATCTCCACCACCATACTGAAC 2659
423 GlnSerLysAspPheCysValMetValGlnIleValProArgValTyrTy 439
2660 CTTACCACTGATTATTGTGTTCTTGTCGAACTCTGGCCAAGAGTCACCTA 2709
439 rHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgP 456
2710 TCATTCCCCCAGCTATGTTTACGGCCTGTTTGAG...AGATCCAACCGAC 2756


456 roLysArgGluProValSerLeuThrLeuAlaValMetLeuGly..... 470
2757 ACAAAAGAGAACCGGTGTCGTTAACCCTGGCCCTATTATTGGGTGGACTA 2806
471 .....LeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIl 485
2807 ACCATGGGGGGAATTGCCGCTGGAATAGGAACAGGGACTACTGCTCTAAT 2856
485 eThrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaM 502
2857 G...GCCACTCAGCAATTCCAG.....CAGCTCCAAGCCGCAG 2891
502 etThrGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGlu 518
2892 TACAGGATGATCTCAGGGAGGTTGAAAAATCAATCTCTAACCTAGAAAAG 2941
519 SerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAs 535
2942 TCTCTCACTTCCCTGTCTGAAGTTGTCCTACAGAATCGAAGGGGCCTAGA 2991
535 pLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluC 552
2992 CTTGTTATTTCTAAAAGAAGGAGGGCTGTGTGCTGCTCTAAAAGAAGAAT 3041
552 ysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLys 568
3042 GTTGCTTCTATGCGGACCACACAGGACTAGTGAGAGACAGCATGGCCAAA 3091
569 LeuArgLysLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGl 585
3092 TTGAGAGAGAGGCTTAATCAGAGACAGAAACTGTTTGAGTCAACTCAAGG 3141
585 yTrpPheGluGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuS 602
3142 ATGGTTTGAGGGACTGTTTAACAGATCCCCTTGGTTTACCACCTTGATAT 3191
602 erAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThrValGly 618
3192 CTACCATTATGGGACCCCTCATTGTACTCCTAATGATTTTGCTCTTCGGA 3241
619 ProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAl 635
3242 CCCTGCATTCTTAATCGATTAGTTCAATTTGTTAAAGACAGGATCTCAGT 3291
635 aValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 648
3292 AGTCCAGGCTTTAGTCCTGACTCAACAATACCACCAGCTA 3331

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 02:58:34 ; Search time 212.72 Seconds
(without alignments)
3147.184 Million cell updates/sec

Title: US-09-171-553B-9
Perfect score: 2956
Sequence: 1 tgctttttagggttaggaac.....aaaaaaaaaaaaaaaaaaaa 2956

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2889.4	97.7	3482	3	US-09-111-085-3	Sequence 3, Appli
	2	1882.8	63.7	7333	4	US-08-766-528-2	Sequence 2, Appli
	3	1563.4	52.9	2462	3	US-09-111-085-1	Sequence 1, Appli
	4	1489.2	50.4	8132	4	US-08-766-528-3	Sequence 3, Appli
	5	1231.8	41.7	8060	4	US-08-766-528-1	Sequence 1, Appli
	6	973.8	32.9	2000	4	US-09-376-781-25	Sequence 25, Appl
	7	973.8	32.9	2000	4	US-09-376-781-30	Sequence 30, Appl
	8	710.4	24.0	1493	4	US-09-376-781-24	Sequence 24, Appl
	9	455	15.4	6312	4	US-09-011-745-7	Sequence 7, Appli
	10	451.2	15.3	8535	3	US-08-716-351A-1	Sequence 1, Appli
	11	372.2	12.6	8655	3	US-09-075-272-1	Sequence 1, Appli
	12	357.8	12.1	10970	3	US-08-716-351A-5	Sequence 5, Appli
	13	343	11.6	698	4	US-09-376-781-18	Sequence 18, Appl
	14	307	10.4	6028	4	US-09-011-745-5	Sequence 5, Appli
	15	303.8	10.3	1965	1	US-08-258-420-9	Sequence 9, Appli
	16	301.4	10.2	2499	1	US-08-105-483-310	Sequence 310, App
	17	301.4	10.2	2499	1	US-08-709-209-310	Sequence 310, App
	18	301.4	10.2	2499	1	US-08-458-101-310	Sequence 310, App
	19	300.6	10.2	2001	3	US-08-850-961-13	Sequence 13, Appl
	20	298.8	10.1	8323	1	US-08-110-300A-8	Sequence 8, Appli
	21	298.8	10.1	8323	2	US-08-886-642-8	Sequence 8, Appli
	22	298.8	10.1	8323	5	PCT-US93-08041-8	Sequence 8, Appli
	23	298.2	10.1	1911	1	US-08-258-420-8	Sequence 8, Appli
c	24	298.2	10.1	8202	1	US-08-258-420-13	Sequence 13, Appl
	25	296.8	10.0	1979	4	US-08-552-369-19	Sequence 19, Appl
	26	293.4	9.9	3925	4	US-09-011-745-9	Sequence 9, Appli
	27	290.2	9.8	1914	1	US-08-258-420-7	Sequence 7, Appli
	28	290	9.8	6061	4	US-09-011-745-6	Sequence 6, Appli
	29	283.6	9.6	8332	3	US-08-850-961-1	Sequence 1, Appli
	30	269.6	9.1	276	4	US-09-376-781-8	Sequence 8, Appli
	31	253.4	8.6	278	4	US-09-376-781-10	Sequence 10, Appl
	32	192	6.5	5109	1	US-08-073-836-1	Sequence 1, Appli
	33	192	6.5	5176	1	US-08-073-836-2	Sequence 2, Appli
	34	157.2	5.3	10306	3	US-08-716-351A-4	Sequence 4, Appli
	35	144.4	4.9	9661	3	US-08-716-351A-3	Sequence 3, Appli
	36	128.2	4.3	445	4	US-09-376-781-7	Sequence 7, Appli
c	37	113	3.8	10367	1	US-08-110-300A-9	Sequence 9, Appli
c	38	113	3.8	10367	2	US-08-886-642-9	Sequence 9, Appli
c	39	113	3.8	10367	5	PCT-US93-08041-9	Sequence 9, Appli
	40	108.2	3.7	1237	1	US-08-798-000-2	Sequence 2, Appli
	41	105.8	3.6	1374	1	US-08-278-630A-9	Sequence 9, Appli
	42	103.8	3.5	1356	2	US-08-484-126-4	Sequence 4, Appli
	43	102.8	3.5	735	1	US-08-798-000-3	Sequence 3, Appli
	44	102.8	3.5	1234	1	US-08-798-000-1	Sequence 1, Appli
	45	102.8	3.5	7308	4	US-09-011-745-3	Sequence 3, Appli

ALIGNMENTS

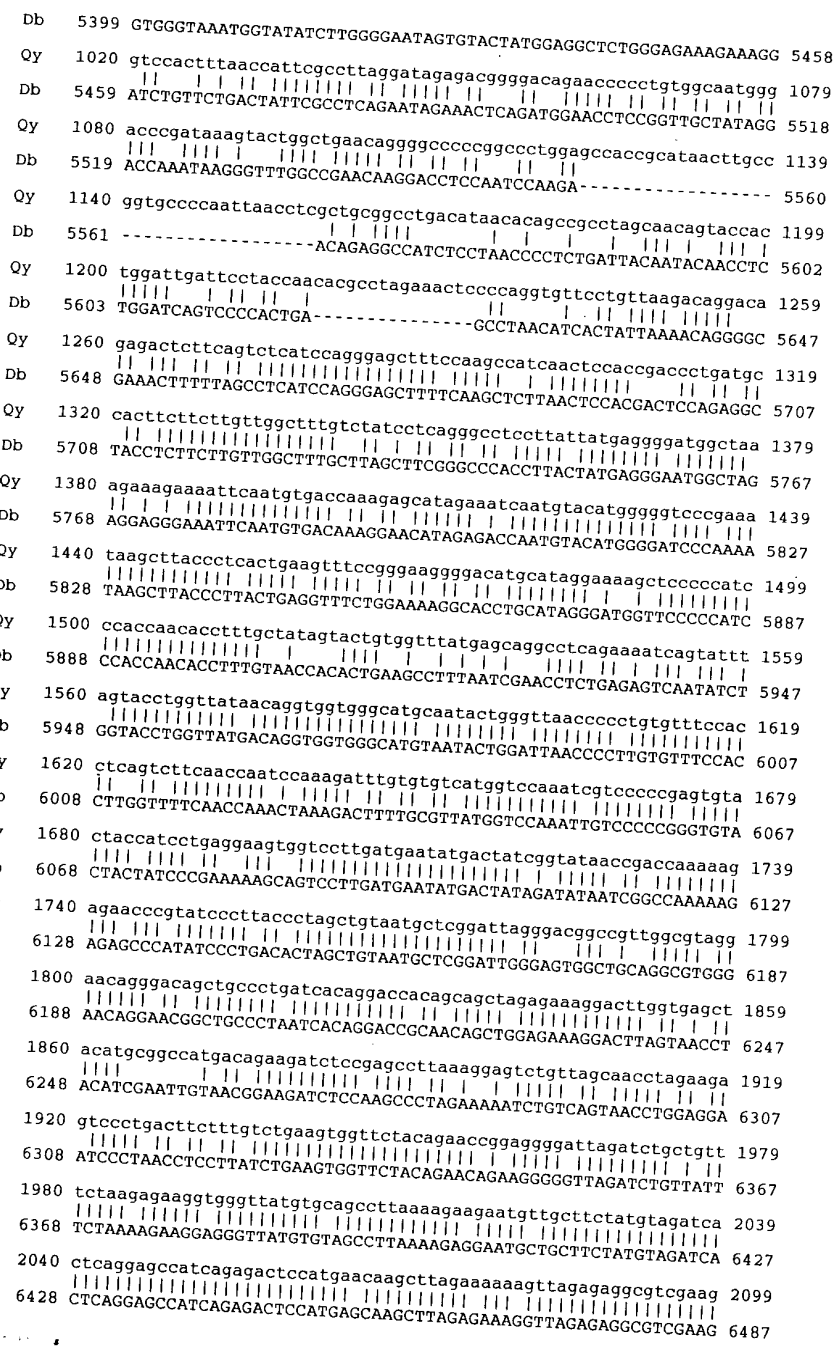
RESULT 1
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

Query Match 97.7%; Score 2889.4; DB 3; Length 3482;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2913; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

```
Qy       1 tgctttttagggttaggaacacccctggacagtttgggctgacccctatgaattgctct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     548 tgctttttagggttaggaacacccctggacagtttgggctgacccctatgaattgctct 607

Qy      61 acgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     608 acgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 667

Qy     121 cccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 180
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     668 cccagcctctgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 727

Qy     181 agctccgggaggcctactcaggagagagagacttgcaagttccacatcgcttccaagttg 240
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||| |
Db     728 agctccgggaggcctactcaggagaaggagacttgcaagttccacatcgcttccaagtgg 787

Qy     241 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggac 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     788 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggcc 847

Qy     301 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     848 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 907
```

```
Qy    361 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    908 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 967

Qy    421 ctgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactccc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    968 ctgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactccc 1027

Qy    481 caggccagtagtaaacgccttatagacagctcgaacccccatagacctttatcccctacc 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   1028 caggccagtagtaaacgccttatagacagctcgaacccccatagacctttatcccttacc 1087

Qy    541 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1088 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 1147

Qy    601 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1148 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 1207

Qy    661 agcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1208 agcacacctcccaacctagtccgtagttatgggttctattgctgcccaggcacagagaaa 1267

Qy    721 gagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacctcc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1268 gagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacctcc 1327

Qy    781 aacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtc 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1328 aacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtc 1387

Qy    841 aattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctccccca 900
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1388 aattccggcccgggcaagtacaaagtgatgaaactatataaagataagagctgctccccca 1447

Qy    901 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1448 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 1507

Qy    961 aagtggataaatggtatgagctgggaatagtttttttaataatatggcgggggagcaggg 1020
          |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db   1508 aagtggataaatggtatgagctgggaatagttttttataaatatggcgggggagcaggg 1567

Qy   1021 tccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcaatggga 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1568 tccactttaaccattcgccttaggatagagacggggacagaaccccctgtggcagtggga 1627

Qy   1081 cccgataaagtactggctgaacaggggccccggccctggagccaccgcataacttgccg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1628 cccgataaagtactggctgaacaggggccccggccctggagccaccgcataacttgccg 1687

Qy   1141 gtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacagtaccact 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   1688 gtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaacggtaccact 1747

Qy   1201 ggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1748 ggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaagacaggacag 1807

Qy   1261 agactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgcc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1808 agactcttcagtctcatccagggagctttccaagccatcaactccaccgaccctgatgcc 1867

Qy   1321 acttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaa 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1868 acttcttcttgttggctttgtctatcctcagggcctccttattatgaggggatggctaaa 1927

Qy   1381 gaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaat 1440
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1928 gaaggaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaat 1987

Qy   1441 aagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1988 aagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcc 2047

Qy   1501 caccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtattta 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2048 caccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagtattta 2107

Qy   1561 gtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2108 gtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttccacc 2167

Qy   1621 tcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtcccccgagtgtac 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2168 tcagtcttcaaccaatccaaagatttctgtgtcatggtccaaatcgtcccccgagtgtac 2227

Qy   1681 taccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaaga 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2228 taccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaaaaga 2287

Qy   1741 gaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtagga 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2288 gaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtagga 2347

Qy   1801 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2348 acagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagcta 2407

Qy   1861 catgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagag 1920
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   2408 catgcggccatgacagaagatctccgagccttagaggagtctgttagcaacctagaagag 2467

Qy   1921 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2468 tccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgttt 2527

Qy   1981 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2528 ctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcac 2587

Qy   2041 tcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagg 2100
          |||||||||||||||||||||||| |||||||||| ||| ||||||||||||||||||||
Db   2588 tcaggagccatcagagactccatgagcaagcttagagaaaggttagagaggcgtcgaagg 2647

Qy   2101 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2648 gaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacc 2707

Qy   2161 accctgctttctgctctgacgggggcccctagtagtcctgctcctgttacttacagttggg 2220
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   2708 accctgctttctgctctgacgggaccccctagtagtcctgctcctgttacttacagttggg 2767

Qy   2221 ccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatc 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2768 ccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatc 2827

Qy   2281 atggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctagcct 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2828 atggtacttaggcaacagtaccaaggccttctgagccaaggagaaactgacctctagcct 2887

Qy   2341 tcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaaaa 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2888 tcccagttctaagattagaactattaacaagacaagaagtggggaatgaaaggatgaaaa 2947

Qy   2401 tgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgcccccgaatt 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2948 tgcaacctaaccctcccagaacccaggaagttaataaaaagctctaaatgcccccgaatt 3007

Qy   2461 acagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggc 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3008 ccagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggc 3067

Qy   2521 ctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggatt 2580
```

```
          ||||||
```



RESULT 3
US-09-111-085-1
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan p
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085

```
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2462
;   TYPE: DNA
;   ORGANISM: Porcine retrovirus
US-09-111-085-1


  Query Match             52.9%;  Score 1563.4;  DB 3;  Length 2462;
  Best Local Similarity   79.5%;  Pred. No. 0;
  Matches 2000;  Conservative    0;  Mismatches  401;  Indels  116;  Gaps    8;

Qy    151 tcgagtgggtgaggcaacgagcgtggaagcagctccgggaggcctactcaggagagagag 210
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||   |||
Db      1 tcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggag---gag 57

Qy    211 acttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtg 270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     58 acttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtg 117

Qy    271 caggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacgg 330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 caggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacgg 177

Qy    331 ctgtgaaagtcgaaggaatctccacctggatccatgcatcccacgttaagctggcgccac 390
          |||||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Db    178 ctgtgaaagtcgaaggaatccccacctggatccatgcatcccacgttaagccggcgccac 237

Qy    391 ctcccgactcggggtggagagccgaaaagactgagaatcccctaagcttcgcctccatc 450
          |||||||| |||||||||| ||||||||||||||| ||||||||||||||||||||||||
Db    238 ctcccgattcggggtggaaagccgaaaagactgaaaatcccctaagcttcgcctccatc 297

Qy    451 gcctggttccttactctaacaataactccccaggccagtagtaaacgccttatagacagc 510
          || |||||||||||||  |||||||||| || |  | | |||||||||| | ||||||
Db    298 gcgtggttccttactctgtcaataactcctcaagttaatggtaaacgccttgtggacagc 357

Qy    511 tcgaacccccatagacctttatcccctacctggctgattattgaccctgatacgggtgtc 570
          ||||| |||||| ||| |||||  |  ||||||  |  ||| |||| | | ||| ||| |
Db    358 ccgaactcccataaacccttatctctcacctggttacttactgactccggtacaggtatt 417

Qy    571 actgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaactgcatttc 630
          | | | || ||||||| ||| |  |||||    || ||||||||||||||  |  || ||
Db    418 aatattaacagcactcaaggggaggctccttggggacctggtggcctgaattatatgtc 477

Qy    631 tgcctccgattgattaaccccgctgtta------aaagcacacctcccaacctagtccgt 684
          ||||| ||||   | | ||| | | |      |  ||||||| ||| |  || |||||
Db    478 tgccttcgatcagtaatccctggtctcaatgaccaggccacacccccgatgtactccgt 537

Qy    685 agttatgggttctattgctgcccagg---cacagagaaagagaaatactgtgggggttct 741
            ||| ||||| ||    ||||||||   | || | || ||  |||| |||||   | ||
Db    538 gcttacgggttttacgtttgcccaggaccccaaataatgaagaatattgtggaaatcct 597

Qy    742 ggggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaatgg 801
            ||| | ||| || | |  ||||||||||| | || || || |||||  | |||||||||
Db    598 caggatttcttttgcaagcaatggagctgcataacttctaatgatgggaattggaaatgg 657

Qy    802 ccgatctctctccaggaccgggtaaaattctcctttgtcaattcc--------------- 846
          ||  ||||||  || ||| | ||||  | ||| ||||| ||
Db    658 ccagtctctcagcaagacagagtaagttactcttttgttaacaatcctaccagttataat 717

Qy    847 ---------------------------------ggcccgggcaagtacaaaatgatgaa 872
                                            |||    ||  |||||||| ||||
Db    718 caatttaattatggccatgggagatggaaagattggcaacagcgggtacaaaaagatgt- 776

Qy    873 actatataaagataagagctgctccccatcagacttagattatctaaagataagtttcac 932
          || | ||||  | |  |||||     | | |||| |||||||  |||| ||||||||||
Db    777 acgaaataagcaaataagctgtcattcgttagacctagattacttaaaaataagtttcac 836

Qy    933 tgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagctggggaatagt 992
          |||||||||||||||| |||||||||||||||| ||||||||||    ||||||||||
Db    837 tgaaaaaggaaaacaagaaaatattcaaaagtgggtaaatggtatatcttggggaatagt 896

Qy    993 ttttta---ataatatggcgggggagcagggtccactttaaccattcgccttaggataga 1049
            |  ||   |   | |||  |      ||| ||   | | || |||||||| || |||||
Db    897 gtactatggaggctctgggagaaagaaaggatctgttctgactattcgcctcagaataga 956

Qy   1050 gacggggacagaaccccctgtggcaatgggacccgataaagtactggctgaacaggggcc 1109
            ||   ||  ||||| || || || || |||||  |||| |   |||| ||||| || ||
Db    957 aactcagatggaacctccggttgctataggaccaaataagggtttggccgaacaaggacc 1016

Qy   1110 cccggccctggagccaccgcataacttgccggtgccccaattaacctcgctgcggcctga 1169
            ||   ||  || |                                          | ||||
Db   1017 tccaatccaagaaca----------------------------------gaggccatc 1040

Qy   1170 cataacacagccgcctagcaacagtaccactggattgattcctaccaacacgcctagaaa 1229
            |  |   |   | ||| |  ||| ||||||   | || ||  |
Db   1041 tcctaacccctctgattacaatacaacctctggatcagtccccactga------------ 1088

Qy   1230 ctccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagggagcttt 1289
                ||     |  || |||| ||||   || ||| || || ||||||||||||||||
Db   1089 ---gcctaacatcactattaaaacaggggcgaaacttttagcctcatccagggagcttt 1145

Qy   1290 ccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtctatcctc 1349
            |||||  | ||||||||    || || || || ||||||||||||||||  || | ||
Db   1146 tcaagctcttaactccacgactccagaggctacctcttcttgttggctttgcttagcttc 1205

Qy   1350 agggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagca 1409
            || || ||||| |||||||| ||||||| || | | ||||||||||||||| || || ||
Db   1206 gggcccaccttactatgagggaatggctagaggagggaaattcaatgtgacaaaggaaca 1265

Qy   1410 tagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagtttccgggaa 1469
            |||| | ||||||||||||||| |||| |||||||||||||||| ||||| ||||| || ||
Db   1266 tagagaccaatgtacatggggatcccaaaataagcttacccttactgaggtttctggaaa 1325

Qy   1470 ggggacatgcataggaaaagctccccccatcccaccaacaccttttgctatagtactgtggt 1529
            || || |||||||| |  | ||||||||||||||||||||||||  |    ||||  |
Db   1326 aggcacctgcatagggatggttcccccatcccaccaacaccttttgtaaccacactgaagc 1385

Qy   1530 ttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggcatg 1589
            | | | |    |||| || | ||| ||| | |||||||||||| ||||||||||||||||
Db   1386 ctttaatcgaacctctgagagtcaatatctggtacctggttatgacaggtggtgggcatg 1445

Qy   1590 caatactgggttaacccccttgtgtttccacctcagtcttcaaccaatccaaagatttgtg 1649
            |||||||| |||||||| |||||||||||||  || |||||||| | ||||| || ||
Db   1446 taatactggattaacccccttgtgtttccacctttggttttcaaccaaactaaagacttttg 1505

Qy   1650 tgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgatga 1709
            || |||||||||| |||||||| |||||||| |||| ||  |||  ||||||||||
Db   1506 cgttatggtccaaattgtcccccgggtgtactactatcccgaaaaagcagtccttgatga 1565

Qy   1710 atatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctagctgtaat 1769
            |||||||||| | ||||| || |||||||||||| ||| ||||||| || ||||||||||
Db   1566 atatgactatagatataatcggccaaaaagagagcccatatccctgacactagctgtaat 1625

Qy   1770 gctcggattagggacggccgttggcgtaggaacagggacagctgccctgatcacaggacc 1829
            |||||||||  ||   ||| |  ||||| |||||||| || |||||||| ||||||||||
Db   1626 gctcggattgggagtggctgcaggcgtgggaacaggaacggctgccctaatcacaggacc 1685

Qy   1830 acagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatctccgagc 1889
            || ||||| |||||||||||| || | ||||||        | || |||||||||| |||
Db   1686 gcaacagctggagaaaggacttagtaacctacatcgaattgtaacggaagatctccaagc 1745

Qy   1890 cttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaagtggttct 1949
            | || |  | ||||| || ||||| || || ||||| || || || ||||||||||||||
Db   1746 cctagaaaaatctgtcagtaacctggaggaatccctaacctccttatctgaagtggttct 1805

Qy   1950 acagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtgcagcctt 2009
            ||||||| | ||||| |||||||||| | ||||||| |||||| |||||||||| ||||||
Db   1806 acagaacagaaggggggttagatctgttatttctaaaagaaggagggttatgtgtagcctt 1865

Qy   2010 aaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactccatgaacaa 2069
          |||||| ||||| ||||||||||||||||||||||||||||||||||||||||||| |||
```

```
Db   1866 aaaagaggaatgctgcttctatgtagatcactcaggagccatcagagactccatgagcaa 1925

Qy   2070 gcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttga 2129

Db   1926 gcttagagaaaggttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttga 1985

Qy   2130 aggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggcccct 2189

Db   1986 aggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggcccct 2045

Qy   2190 agtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgcctt 2249

Db   2046 agtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgcctt 2105

Qy   2250 tgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggcct 2309

Db   2106 tgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggcct 2165

Qy   2310 tctgagccaaggagaaactgacctctagccttcccagttctaagattagaactattaaca 2369

Db   2166 tctgagccaaggagaaactgacctctagccttcccagttctaagattagaactattaaca 2225

Qy   2370 agacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaa 2429

Db   2226 agacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaa 2285

Qy   2430 gttaataaaaagctctaaatgcccccgaattacagaccctgctggctgccagtaaatagg 2489

Db   2286 gttaataaaaagctctaaatgcccccgaattccagaccctgctggctgccagtaaatagg 2345

Qy   2490 tagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaagataacag 2549

Db   2346 tagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaagataacag 2405

Qy   2550 gaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccatagaa 2606

Db   2406 gaaatgagttgactaatcgcttatctggattctgtaaaaccgactggcaccatagaa 2462
```

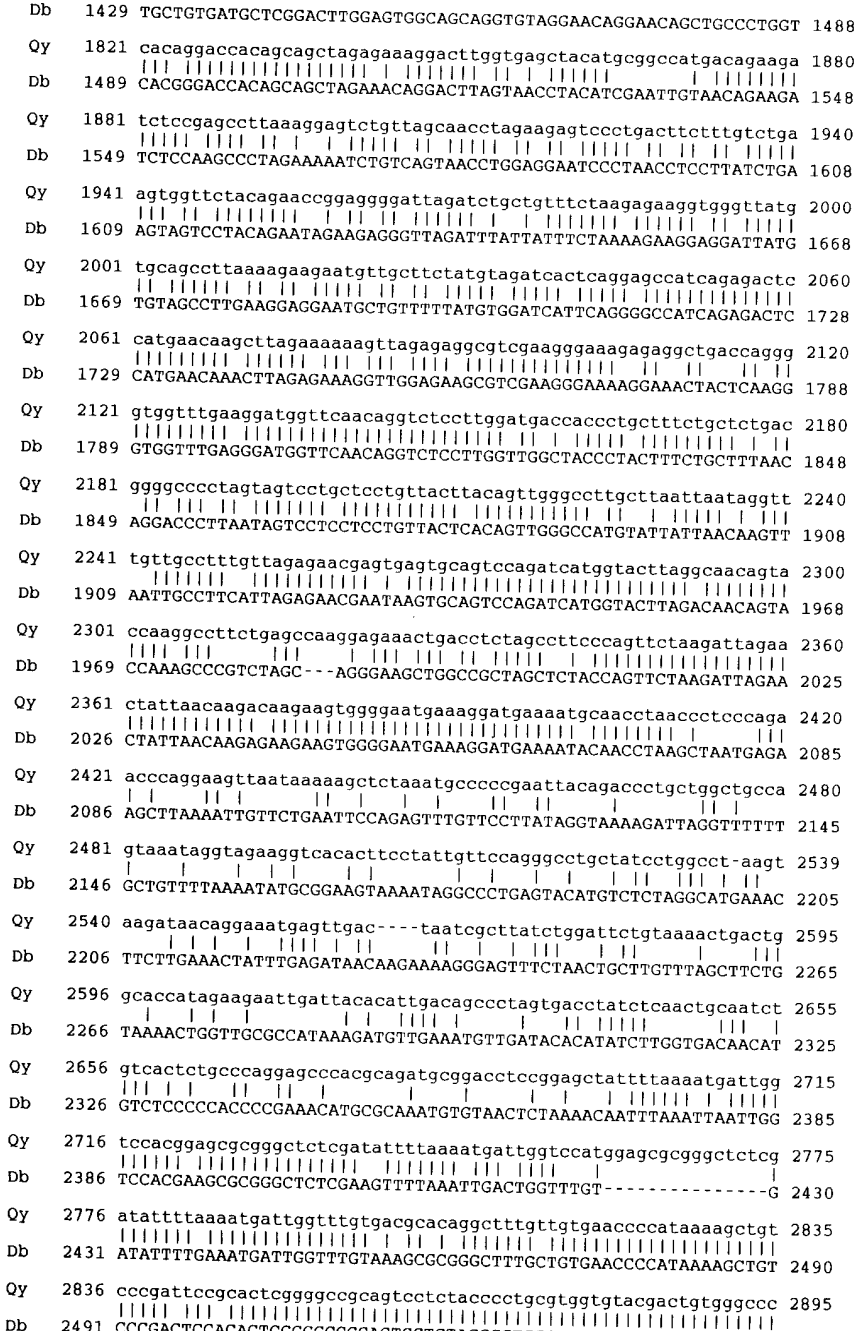
```
RESULT   4
US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
```

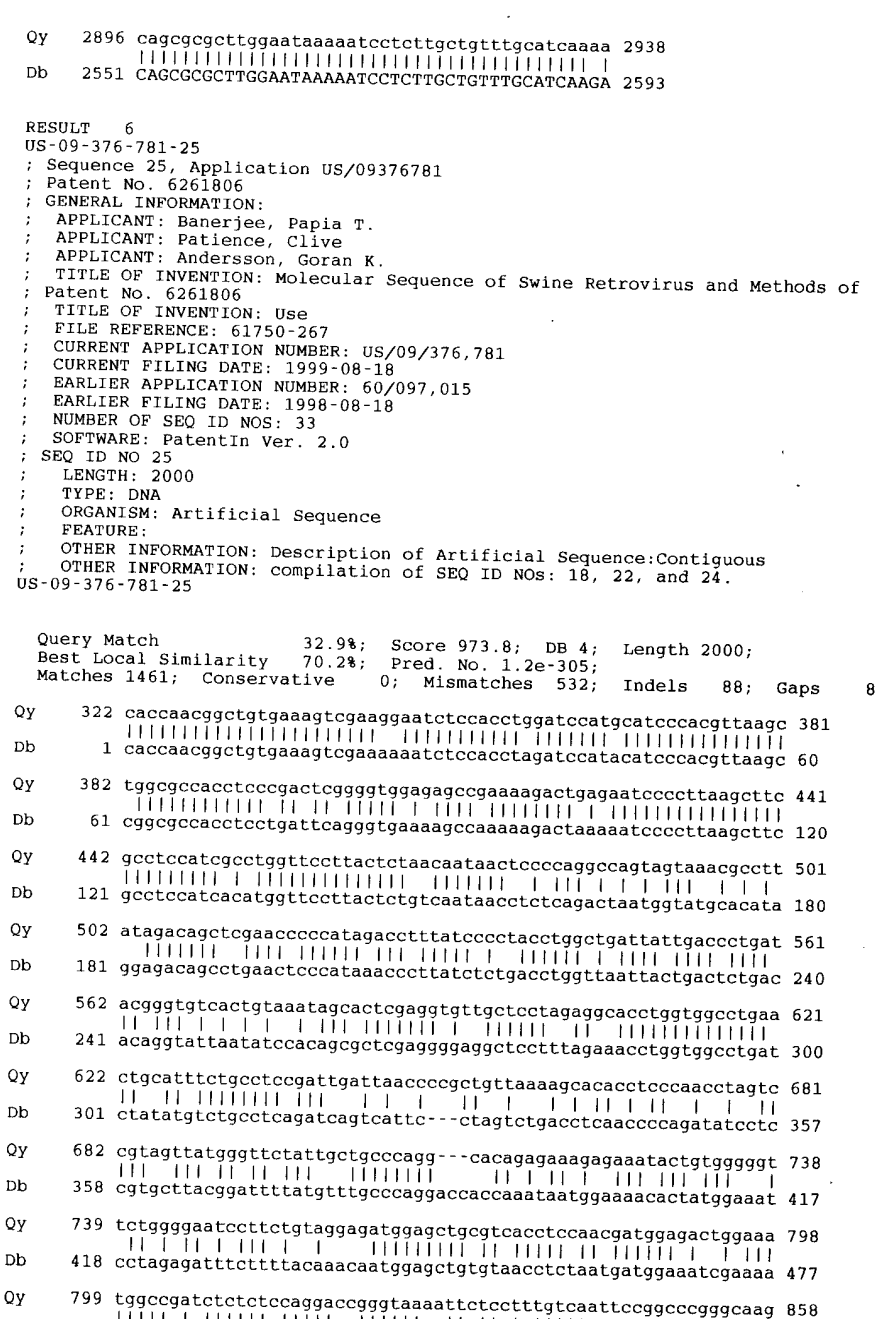
```
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8132 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-3


  Query Match            50.4%;  Score 1489.2;  DB 4;  Length 8132;
  Best Local Similarity  72.2%;  Pred. No. 0;
  Matches 2140;  Conservative    0;  Mismatches  713;  Indels  113;  Gaps   11;

Qy      1 tgctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgctct 60

Db   5257 TGCTTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCCTATGAATTACTCT 5316

Qy     61 acgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgcttt 120

Db   5317 ACGGGGGACCCCCCCCATTGGTAGAAATTGCTTCTGTACATAGTGCTGACGTGCTGCTTT 5376

Qy    121 cccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagc 180

Db   5377 CCCAGCCTTTGTTCTCTAGGCTCAAGGCACTTGAGTGGGTGAGACAACGAGCGTGGAGGC 5436

Qy    181 agctccgggaggcctactcaggagagagagacttgcaagttccacatcgcttccaagttg 240

Db   5437 AACTCCGGGAGGCCTACTCAGGAGGAGGAGACTTGCAGATCCCACATCGTTTCCAAGTGG 5496

Qy    241 gagattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggac 300

Db   5497 GAGATTCAGTCTACGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGCC 5556

Qy    301 cttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctgga 360

Db   5557 CTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGA 5616

Qy    361 tccatgcatcccacgttaagctggcgccacctcccgactcggggtggagagccgaaaaga 420

Db   5617 TCCATGCATCCCACGTTAAACCGGCGCCACCTCCCGATTCGGGGTGGAAAGCCGAAAAGA 5676

Qy    421 ctgagaatccccttaagcttcgcctccatcgcctggttccttactctaacaataactccc 480

Db   5677 CTGAAAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTCAATAACCTCT 5736

Qy    481 caggccagtagtaaacgccttatagacagctcgaaccccccatagacctttatcccctacc 540

Db   5737 CAGACTAATGGTATGCGCATAGGAGACAGCCTGAACTCCCATAAACCCTTATCTCTCACC 5796

Qy    541 tggctgattattgaccctgatacgggtgtcactgtaaatagcactcgaggtgttgctcct 600

Db   5797 TGGTTAATTACTGACTCCGGCACAGGTATTAATATCAACAACACTCAAGGGGAGGCTCCT 5856

Qy    601 agaggcacctggtggcctgaactgcatttctgcctccgattgattaaccccgctgttaaa 660

Db   5857 TTAGGAACCTGGTGGCCTGATCTATACGTTTGCCTCAGATCAGTTATTC---CTAGTCTG 5913

Qy    661 agcacacctcccaacctagtccgtagttatgggttctattgctgcccagg---cacagag 717

Db   5914 ACCTCACCCCCAGATATCCTCCATGCTCACGGATTTTATGTTTGCCCAGGACCACCAAAT 5973

Qy    718 aaagagaaatactgtgggggttctggggaatccttctgtaggagatggagctgcgtcacc 777

Db   5974 AATGGAAAACATTGCGGAAATCCCAGAGATTTCTTTTGTAAACAATGGAACTGTGTAACC 6033

Qy    778 tccaacgatggagactggaaatggccgatctctctccaggaccgggtaaaattctccttt 837

Db   6034 TCTAATGATGGATATTGGAAATGGCCAACCTCTCAGCAGGATAGGGTAAGTTTTTCTTAT 6093

Qy    838 gtcaa------------ttccggcccgggcaagtacaaaatgatgaaactatataaagat 885

Db   6094 GTCAACACCTATACCAGCTCTGGACAATTTAATTACCTGACCTGGATTAGAACTGGAAGC 6153
```

```
Qy    886 aagagctgctccccatcagacttagattatctaaagataagtttcactgaaaaaggaaaa 945
Db   6154 CCCAAGTGCTCTCCTTCAGACCTAGATTACCTAAAAATAAGTTTCACTGAGAAAGGAAAA 6213

Qy    946 caggaaaatattcaaaagtggataaatggtatgagctggggaatagtttttta---ataa 1002
Db   6214 CAAGAAAATATCCTAAAATGGGTAAATGGTATGTCTTGGGGAATGGTATATTATGGAGGC 6273

Qy   1003 tatggcgggggagcagggtccacttaaccattcgccttaggatagagacggggacagaa 1062
Db   6274 TCGGGTAAACAACCAGGCTCCATTCTAACTATTCGCCTCAAAATAAACCAGCTG---GAG 6330

Qy   1063 ccccctgtggcaatgggacccgataaagtactggctgaacaggggccccggccctggag 1122
Db   6331 CCTCCAATGGCTATAGGACCAAATACGGTCTTGACGGGTCAAAGACCCCCAACCC----- 6385

Qy   1123 ccaccgcataacttgccggtgccccaattaacctcgctgcggcctgacataacacagccg 1182
Db   6386 ----------------------------------------------AAGGACCAGGACCA 6399

Qy   1183 cctagcaacagtaccactggattgattcctaccaacacgcctagaaactccccaggtgtt 1242
Db   6400 TCCTCTAACATAACTTCTGGATCAGACCCCACTGAGTCTAACAGCA-------------- 6445

Qy   1243 cctgttaagacaggacagagactcttcagtctcatccagggagctttccaagccatcaac 1302
Db   6446 -CGACTAAAATGGGGGCAAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAAC 6504

Qy   1303 tccaccgaccctgatgccacttcttcttgttggctttgtctatcctcagggcctccttat 1362
Db   6505 TCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTATGCTTAGCTTCGGGCCCACCTTAC 6564

Qy   1363 tatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgt 1422
Db   6565 TATGAAGGAATGGCTAGAAGAGGGAAATTCAATGTGACAAAAGAACATAGAGACCAATGC 6624

Qy   1423 acatgggggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgcata 1482
Db   6625 ACATGGGGATCCCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATA 6684

Qy   1483 ggaaaagctccccccatcccaccaacaccttttgctatagtactgtggtttatgagcaggcc 1542
Db   6685 GGAAAGGTTCCCCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCAAACC 6744

Qy   1543 tcagaaaatcagtatttagtacctggttataacaggtggtgggcatgcaatactgggtta 1602
Db   6745 TCTGAGAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTA 6804

Qy   1603 acccccctgtgtttccacctcagtcttcaaccaatccaaagatttgtgtgtcatggtccaa 1662
Db   6805 ACCCCTTGTGTTTCCACCTTGGTTTTTAACCAAACTAAAGATTTTTGCATTATGGTCCAA 6864

Qy   1663 atcgtcccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcgg 1722
Db   6865 ATTGTTCCCCGAGTGTATTACTATCCCGAAAAAGCAATCCTTGATGAATATGACTACAGA 6924

Qy   1723 tataaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggattaggg 1782
Db   6925 AATCATCGACAAAAGAGAGAACCCATATCTCTGACACTTGCTGTGATGCTCGGACTTGGA 6984

Qy   1783 acggccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagag 1842
Db   6985 GTGGCAGCAGGTGTAGGAACAGGAACAGCTGCCCTGGTCACGGGACCACAGCAGCTAGAA 7044

Qy   1843 aaaggacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtct 1902
Db   7045 ACAGGACTTAGTAACCTACATCGAATTGTAACAGAAGATCTCCAAGCCCTAGAAAAATCT 7104

Qy   1903 gttagcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggagg 1962
Db   7105 GTCAGTAACCTGGAGGAATCCCTAACCTCCTTATCTGAAGTAGTCCTACAGAATAGAAGA 7164

Qy   1963 ggattagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgt 2022
Db   7165 GGGTTAGATTTATTATTTCTAAAAGAAGGAGGATTATGTGTAGCCTTGAAGGAGGAATGC 7224

Qy   2023 tgcttctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaag 2082
Db   7225 TGTTTTTATGTGGATCATTCAGGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAAGG 7284

Qy   2083 ttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaac 2142
Db   7285 TTGGAGAAGCGTCGAAGGGAAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGGTTCAAC 7344

Qy   2143 aggtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctc 2202
Db   7345 AGGTCTCTTTGGTTGGCTACCCTACTTTCTGCTTTAACAGGACCCTTAATAGTCCTCCTC 7404

Qy   2203 ctgttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacga 2262
Db   7405 CTGTTACTCACAGTTGGGCCATGTATTATTAACAAGTTAATTGCCTTCATTAGAGAACGA 7464

Qy   2263 gtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaagga 2322
Db   7465 ATAAGTGCAGTCCAGATCATGGTACTTAGACAACAGTACCAAAGCCCGTCTAGC---AGG 7521

Qy   2323 gaaactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtgg 2382
Db   7522 GAAGCTGGCCGCTAGCTCTACCAGTTCTAAGATTAGAACTATTAACAAGAGAAGAAGTGG 7581

Qy   2383 ggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagttaataaaaagc 2442
Db   7582 GGAATGAAAGGATGAAAATACAACCTAAGCTAATGAGAAGCTTAAAATTGTTCTGAATTC 7641

Qy   2443 tctaaatgccccccgaattacagaccctgctggctgccagtaaataggtagaaggtcacac 2502
Db   7642 CAGAGTTTGTTCCTTATAGGTAAAAGATTAGGTTTTTTGCTGTTTTAAAATATGCGGAAG 7701

Qy   2503 ttcctattgttccagggcctgctatcctggcct-aagtaagataacaggaaatgagttga 2561
Db   7702 TAAAATAGGCCCTGAGTACATGTCTCTAGGCATGAAACTTCTTGAAACTATTTGAGATAA 7761

Qy   2562 c----taatcgcttatctggattctgtaaaactgactggcaccatagaagaattgattac 2617
Db   7762 CAAGAAAAGGGAGTTTCTAACTGCTTGTTTAGCTTCTGTAAAACTGGTTGCGCCATAAAG 7821

Qy   2618 acattgacagccctagtgacctatctcaactgcaatctgtcactctgcccaggagcccac 2677
Db   7822 ATGTTGAAATGTTGATACACATATCTTGGTGACAACATGTCTCCCCCACCCCGAAACATG 7881

Qy   2678 gcagatgcggacctccggagctattttaaaatgattggtccacggagcgcgggctctcga 2737
Db   7882 CGCAAATGTGTAACTCTAAAACAATTTAAATTAATTGGTCCACGAAGCGCGGGCTCTCGA 7941

Qy   2738 tattttaaaatgattggtccatggagcgcgggctctcgatattttaaaatgattggtttg 2797
Db   7942 AGTTTTAAATTGACTGGTTTGT--------------GATATTTTGAAATGATTGGTTTG 7986

Qy   2798 tgacgcacaggctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccg 2857
Db   7987 TAAAGCGCGGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGACTCCACACTCGGGGCCG 8046

Qy   2858 cagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatc 2917
Db   8047 CAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATC 8106

Qy   2918 ctcttgctgtttgcatcaaaaaaaaa 2943
Db   8107 CTCTTGCTGTTTGCATCAAAAAAAAA 8132
```

RESULT 5
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861

```
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1

  Query Match             41.7%;  Score 1231.8;  DB 4;  Length 8060;
  Best Local Similarity   69.8%;  Pred. No. 0;
  Matches 1873;  Conservative    0;  Mismatches  697;  Indels  113;  Gaps   11;

Qy    279 ctcgagactcggtggaagggaccttatctcgtacttttgaccacaccaacggctgtgaaa 338
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      1 CTCGAGACTCGGTGGAAGGGCCCTTATCTCGTACTTTTGACCACACCAACGGCTGTGAAA 60

Qy    339 gtcgaaggaatctccacctggatccatgcatcccacgttaagctggcgccacctcccgac 398
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     61 GTCGAAGGAATCTCCACCTGGATCCATGCATCCCACGTTAAGCCGGCGCCACCTCCCGAT 120

Qy    399 tcggggtggagagccgaaaagactgagaatcccctaagcttcgcctccatcgcctggtt 458
          |||||||||| |||||||||||||| |||||||||||||||||||||||||||| |||||
Db    121 TCGGGGTGGAAAGCCGAAAAGACTGAAAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTT 180

Qy    459 ccttactctaacaataactccccaggccagtagtaaacgccttatagacagctcgaaccc 518
          |||||||||  |||||||  | ||| | | | |||  ||| |   |||||||  |||| |
Db    181 CCTTACTCTGTCAATAACCTCTCAGACTAATGGTATGCGCATAGGAGACAGCCTGAACTC 240

Qy    519 ccatagacctttatcccctacctggctgattattgaccctgatacgggtgtcactgtaaa 578
          |||||  ||| |||||  |  |||||| | |||| |||| | |  || ||| | | | | ||
Db    241 CCATAAACCCTTATCTCTCACCTGGTTAATTACTGACTCCGGCACAGGTATTAATATCAA 300

Qy    579 tagcactcgaggtgttgctcctagaggcacctggtggcctgaactgcatttctgcctccg 638
          | |||||  ||| |  |||||  ||| |||||||||||||| ||  |  | ||||||  |
Db    301 CAACACTCAAGGGGAGGCTCCTTTAGGAACCTGGTGGCCTGATCTATACGTTTGCCTCAG 360

Qy    639 attgattaaccccgctgttaaaagcacacctcccaacctagtccgtagttatgggttcta 698
          ||   |||  |   ||  |   | | |||| ||  |  |  |||  |  | | || || ||
Db    361 ATCAGTTATTC---CTAGTCTGACCTCACCCCCAGATATCCTCCATGCTCACGGATTTTA 417

Qy    699 ttgctgcccagg---cacagagaaagagaaatactgtgggggttctggggaatccttctg 755
          |   ||||||||     || | || |  ||| | || ||   | |  | || | ||| ||
Db    418 TGTTTGCCCAGGACCACCAAATAATGGAAAACATTGCGGAAATCCCAGAGATTTCTTTTG 477

Qy    756 taggagatggagctgcgtcacctccaacgatggagactggaaatggccgatctctctcca 815
          ||    ||||| ||| || ||||| || |||||| | |||||||||||| | |||||  ||
Db    478 TAAACAATGGAACTGTGTAACCTCTAATGATGGATATTGGAAATGGCCAACCTCTCAGCA 537

Qy    816 ggaccgggtaaaattctccttttgtcaa-----------ttccggcccgggcaagtacaa 863
          |||  ||||||  || || | ||||||              || || |     || |||
Db    538 GGATAGGGTAAGTTTTTCTTATGTCAACACCTATACCAGCTCTGGACAATTTAATTACCT 597

Qy    864 aatgatgaaactatataaagataagagctgctccccatcagacttagattatctaaagat 923
          |   ||    |  | |       |  ||||| || |||||| ||||||| ||||| ||
Db    598 GACCTGGATTAGAACTGGAAGCCCCAAGTGCTCTCCTTCAGACCTAGATTACCTAAAAAT 657

Qy    924 aagtttcactgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatgagctg 983
          |||||||||||| |||||||||| |||||||| | ||| ||| ||||||||||||   ||
Db    658 AAGTTTCACTGAGAAAGGAAAACAAGAAAATATCCTAAAATGGGTAAATGGTATGTCTTG 717

Qy    984 gggaatagtttttta---ataatatggcgggggagcagggtccactttaaccattcgcct 1040
          |||||| || | |||   |   |  ||       | |||| |||| | |||| ||||||||
Db    718 GGGAATGGTATATTATGGAGGCTCGGGTAAACAACCAGGCTCCATTCTAACTATTCGCCT 777

Qy   1041 taggatagagacggggacagaaccccctgtggcaatgggacccgataaagtactggctga 1100
          |  ||| |   |  |   || || ||  |||| || |||||  |||  ||  || | |
Db    778 CAAAATAAACCAGCTG---GAGCCTCCAATGGCTATAGGACCAAATACGGTCTTGACGGG 834

Qy   1101 acaggggcccccggccctggagccaccgcataacttgccggtgccccaattaacctcgct 1160
          ||  | |||||  |||  |  |||   |    ||
Db    835 TCAAAGACCCCCAACCCAAGGACCAGGACCATCCTCT---------------------- 871

Qy   1161 gcggcctgacataacacagccgcctagcaacagtaccactggattgattcctaccaacac 1220
                                                ||||  ||  ||||||     || ||  |  |
Db    872 ----------------------------AACATAACTTCTGGATCAGACCCCACTGAGTC 903

Qy   1221 gcctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctcatcca 1280
             || |               |  ||| |  ||    | ||| || || ||||||||
Db    904 TAGCAGCA--------------CGACTAAAATGGGGGCAAAACTTTTTAGCCTCATCCA 948

Qy   1281 gggagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttg 1340
          ||||||||| |||||  | ||||||||    || || || || ||||||||||||||  ||
Db    949 GGGAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTATG 1008

Qy   1341 tctatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgac 1400
          || | |  || || ||||| |||||| || |||||||  |  | | ||||||||||||||
Db    1009 CTTAGCTTTGGGCCCACCTTACTATGAAGGAATGGCTAGAAGAGGGAAATTCAATGTGAC 1068

Qy   1401 caaagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagt 1460
          ||||| |||||| | ||||| |||||||| |||| ||||||||||||||| ||||| ||
Db    1069 AAAAGAACATAGAGACCAATGCACATGGGGATCCCAAAATAAGCTTACCCTTACTGAGGT 1128

Qy   1461 ttccgggaaggggacatgcataggaaaagctcccccatcccaccaacacctttgctatag 1520
          ||| || || || || |||||||||||  | ||||||||||||||||||||||||||  |
Db    1129 TTCTGGAAAAGGCACCTGCATAGGAAAGGTTCCCCCATCCCACCAACACCTTTGTAACCA 1188

Qy   1521 tactgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtg 1580
          ||||  |  | | | ||  |||| |||| ||| ||| | ||||||||||||| |||||||
Db    1189 CACTGAAGCCTTTAATCAAACCTCTGAAAGTCAATATCTGGTACCTGGTTATGACAGGTG 1248

Qy   1581 gtgggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaa 1640
          |||||||| |||||||| |||||||| ||||||||||||  || || ||||| | ||
Db    1249 GTGGGCATGTAATACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTTAACCAAACTAA 1308

Qy   1641 agatttgtgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggt 1700
          ||||||  ||  | |||||||||||  || |||||||||| ||| |||| ||  |||   |
Db    1309 AGATTTTTGCATTATGGTCCAAATTGTTCCCCGAGTGTATTACTATCCCGAAAAAGCAAT 1368

Qy   1701 ccttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccct 1760
          ||||||||||||||||||  |  || | |||| ||| ||||||||| |||| || || ||
Db    1369 CCTTGATGAATATGACTACAGAAATCATCGACAAAAGAGAGAACCCATATCTCTGACACT 1428

Qy   1761 agctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgat 1820
          |||||  |||||||||  | ||   |||  |  ||||||||||| ||||||||||| |
```





RESULT 6
US-09-376-781-25
; Sequence 25, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Contiguous
; OTHER INFORMATION: compilation of SEQ ID NOs: 18, 22, and 24.
US-09-376-781-25

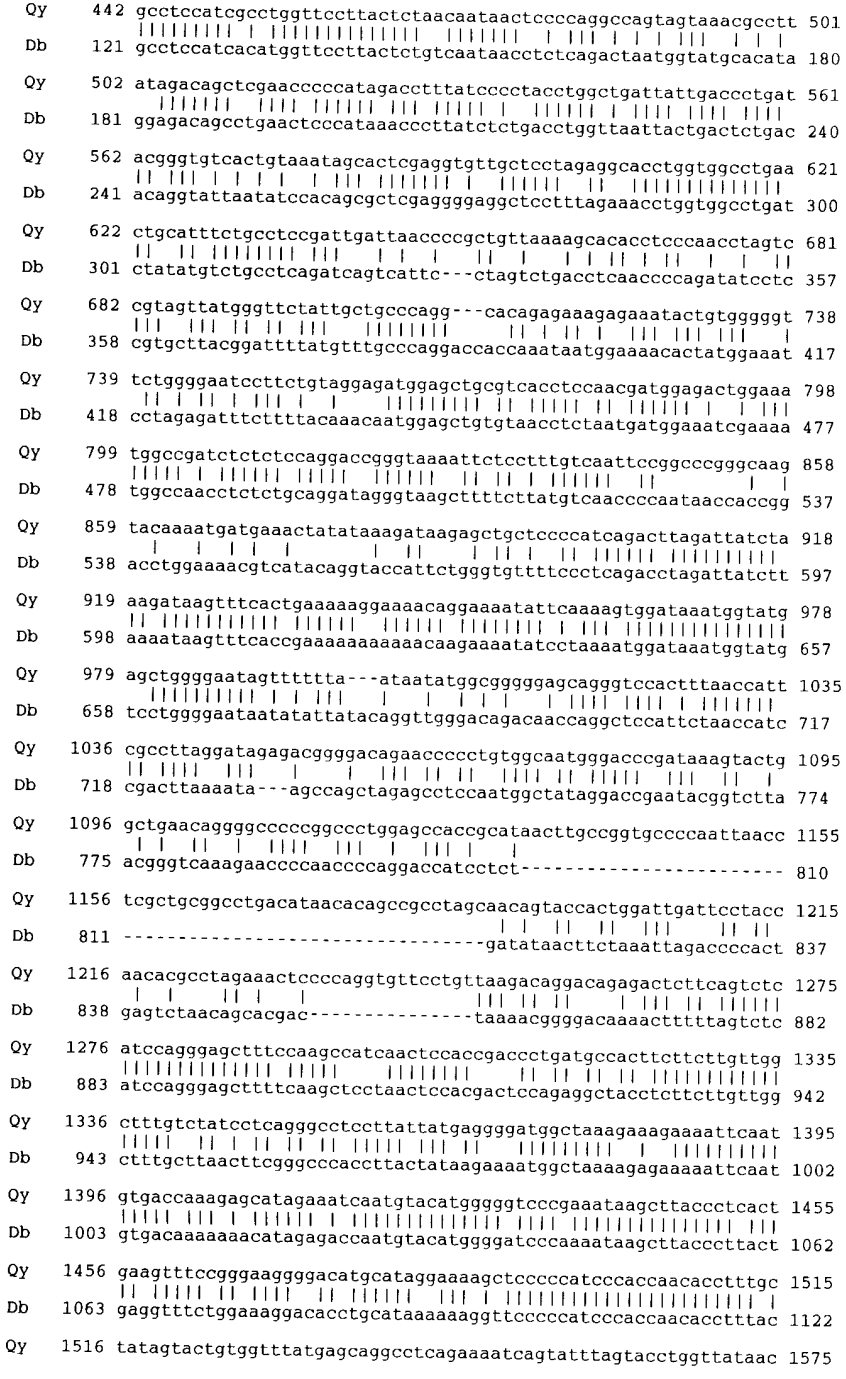
Query Match 32.9%; Score 973.8; DB 4; Length 2000;
Best Local Similarity 70.2%; Pred. No. 1.2e-305;
Matches 1461; Conservative 0; Mismatches 532; Indels 88; Gaps 8;

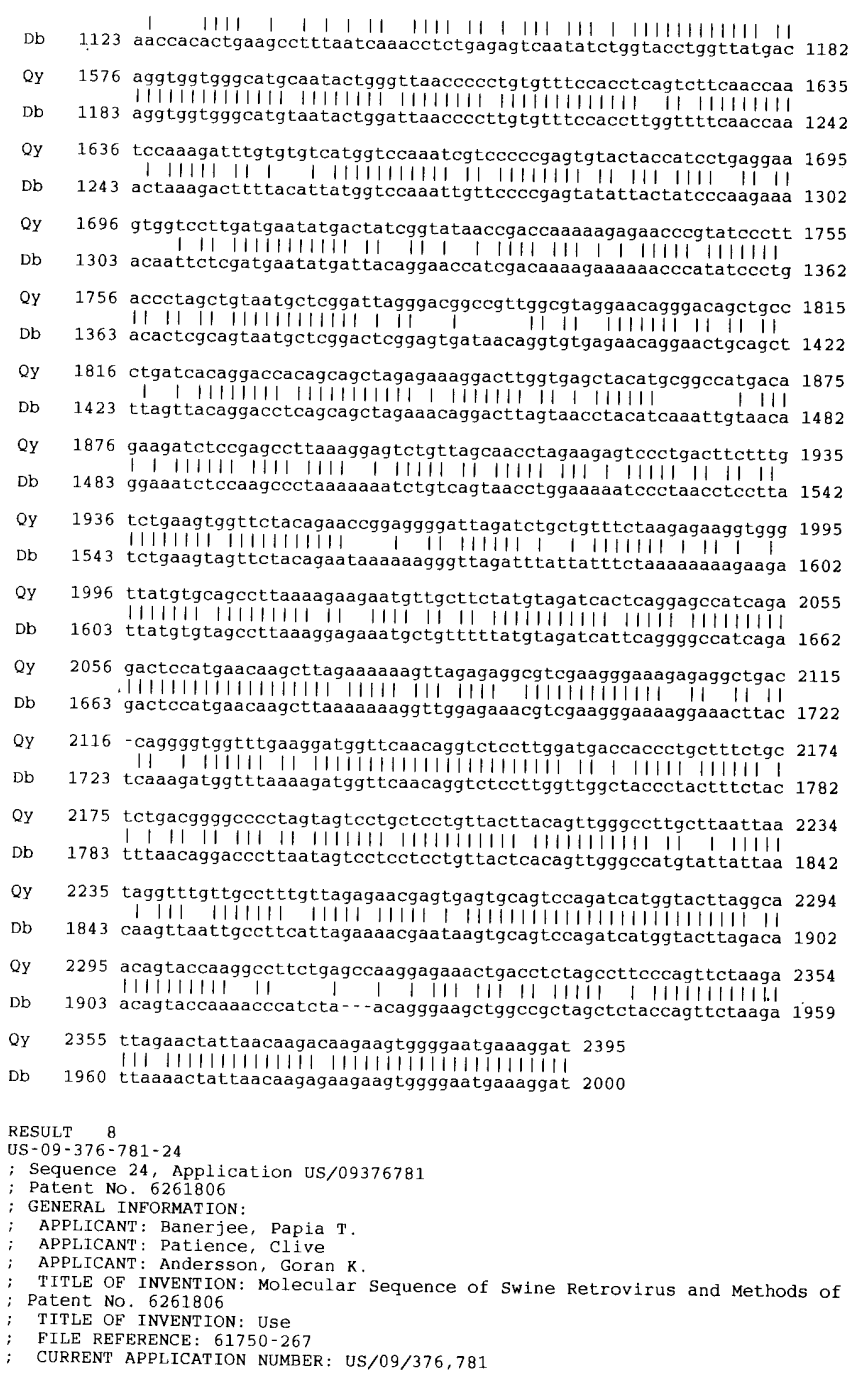
```
Qy    859 tacaaaatgatgaaactatataaagataagagctgctccccatcagacttagattatcta 918
Db    538 acctggaaaacgtcatacaggtaccattctgggtgttttccctcagacctagattatctt 597

Qy    919 aagataagtttcactgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatg 978
Db    598 aaaataagtttcaccgaaaaaaaaaaacaagaaaatatcctaaaatggataaatggtatg 657

Qy    979 agctggggaatagtttttta---ataatatggcgggggagcagggtccactttaaccatt 1035
Db    658 tcctggggaatataatattatacaggttgggacagacaaccaggctccattctaaccatc 717

Qy   1036 cgccttaggatagagacggggacagaacccccctgtggcaatgggacccgataaagtactg 1095
Db    718 cgacttaaaata---agccagctagagcctccaatggctataggaccgaatacggtctta 774

Qy   1096 gctgaacaggggccccccggccctggagccaccgcataacttgccggtgccccaattaacc 1155
Db    775 acgggtcaaagaaccccaaccccaggaccatcctct---------------------- 810

Qy   1156 tcgctgcggcctgacataacacagccgcctagcaacagtaccactggattgattcctacc 1215
Db    811 --------------------------------gatataacttctaaattagaccccact 837

Qy   1216 aacacgcctagaaactccccaggtgttcctgttaagacaggacagagactcttcagtctc 1275
Db    838 gagtctaacagcacgac--------------taaaacggggacaaaacttttagtctc 882

Qy   1276 atccagggagctttccaagccatcaactccaccgaccctgatgccacttcttcttgttgg 1335
Db    883 atccagggagcttttcaagctcctaactccacgactccagaggctacctcttcttgttgg 942

Qy   1336 ctttgtctatcctcagggcctccttattatgaggggatggctaaagaaagaaaattcaat 1395
Db    943 ctttgcttaacttcgggcccaccttactataagaaaatggctaaaagagaaaaattcaat 1002

Qy   1396 gtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaataagcttaccctcact 1455
Db   1003 gtgacaaaaaaacatagagaccaatgtacatggggatcccaaaataagcttacccttact 1062

Qy   1456 gaagtttccgggaaggggacatgcataggaaaagctcccccatcccaccaacacctttgc 1515
Db   1063 gaggtttctggaaaggacacctgcataaaaaaggttcccccatcccaccaacacctttac 1122

Qy   1516 tatagtactgtggtttatgagcaggcctcagaaaatcagtatttagtacctggttataac 1575
Db   1123 aaccacactgaagcctttaatcaaacctctgagagtcaatatctggtacctggttatgac 1182

Qy   1576 aggtggtgggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaa 1635
Db   1183 aggtggtgggcatgtaatactggattaacccctttgtgtttccaccttggttttcaaccaa 1242

Qy   1636 tccaaagatttgtgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaa 1695
Db   1243 actaaagacttttacattatggtccaaattgttccccgagtatattactatcccaagaaa 1302

Qy   1696 gtggtccttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatccctt 1755
Db   1303 acaattctcgatgaatatgattacaggaaccatcgacaaaagaaaaaacccatatccctg 1362

Qy   1756 accctagctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgcc 1815
Db   1363 acactcgcagtaatgctcggactcggagtgataacaggtgtgagaacaggaactgcagct 1422

Qy   1816 ctgatcacaggaccacagcagctagagaaaggacttggtgagctacatgcggccatgaca 1875
Db   1423 ttagttacaggacctcagcagctagaaacaggacttagtaacctacatcaaattgtaaca 1482

Qy   1876 gaagatctccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttg 1935
Db   1483 ggaaatctccaagccctaaaaaaatctgtcagtaacctggaaaaatccctaacctcctta 1542

Qy   1936 tctgaagtggttctacagaaccggaggggattagatctgctgtttctaagagaaggtggg 1995
Db   1543 tctgaagtagttctacagaataaaaaaggttagatttattatttctaaaaaaaagaaga 1602

Qy   1996 ttatgtgcagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcaga 2055
Db   1603 ttatgtgtagccttaaaggagaaatgctgtttttatgtagatcattcaggggccatcaga 1662

Qy   2056 gactccatgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgac 2115
Db   1663 gactccatgaacaagcttaaaaaaaggttggagaaacgtcgaagggaaaaggaaacttac 1722

Qy   2116 -caggggtggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgc 2174
Db   1723 tcaaagatggtttaaaagatggttcaacaggtctccttggttggctaccctactttctac 1782

Qy   2175 tctgacggggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaa 2234
Db   1783 tttaacaggaccttaatagtcctcctcctgttactcacagttgggccatgtattattaa 1842

Qy   2235 taggtttgttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggca 2294
Db   1843 caagttaattgccttcattagaaaacgaataagtgcagtccagatcatggtacttagaca 1902

Qy   2295 acagtaccaaggccttctgagccaaggagaaactgacctctagccttcccagttctaaga 2354
Db   1903 acagtaccaaaacccatcta---acagggaagctggccgctagctctaccagttctaaga 1959

Qy   2355 ttagaactattaacaagacaagaagtggggaatgaaaggat 2395
Db   1960 ttaaaactattaacaagagaagaagtggggaatgaaaggat 2000
```

RESULT 7
US-09-376-781-30
; Sequence 30, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
; OTHER INFORMATION: consensus sequence of PERV-D.
US-09-376-781-30

Query Match 32.9%; Score 973.8; DB 4; Length 2000;
Best Local Similarity 70.2%; Pred. No. 1.2e-305;
Matches 1461; Conservative 0; Mismatches 532; Indels 88; Gaps 8;

```
Qy    322 caccaacggctgtgaaagtcgaaggaatctccacctggatccatgcatcccacgttaagc 381
Db      1 caccaacggctgtgaaagtcgaaaaaatctccacctagatccatacatcccacgttaagc 60

Qy    382 tggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttc 441
Db     61 cggcgccacctcctgattcagggtgaaaagccaaaaagactaaaaatccccttaagcttc 120
```





RESULT 8
US-09-376-781-24
; Sequence 24, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781

```
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
;   LENGTH: 1493
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: sequence of the 3' end of the PERV-D env region.
US-09-376-781-24


  Query Match             24.0%;  Score 710.4;  DB 4;  Length 1493;
  Best Local Similarity   70.7%;  Pred. No. 3.4e-220;
  Matches 1069;  Conservative    0;  Mismatches  361;  Indels   82;  Gaps    6;

Qy    890 gctgctcccccatcagacttagattatctaaagataagtttcactgaaaaaggaaaacagg 949
          | || |  || |||||| |||||||||| || ||||||||||| ||||||  |||||| |
Db     50 ggtgttttccctcagacctagattatcttaaaataagtttcaccgaaaaaaaaaaacaag 109

Qy    950 aaaatattcaaaagtggataaatggtatgagctggggaatagtttttta---ataatatg 1006
          ||||||| | ||| ||||||||||||||  ||||||||||| | | |||   |   | |
Db    110 aaaatatcctaaaatggataaatggtatgtcctggggaataatatattatacaggttggg 169

Qy   1007 gcgggggagcagggtccactttaaccattcgccttaggatagagacggggacagaacccc 1066
          | |   | |||| |||| | ||||||| || ||||  |||   |     |  ||| || |
Db    170 acagacaaccaggctccattctaaccatccgacttaaaata---agccagctagagcctc 226

Qy   1067 ctgtggcaatgggacccgataaagtactggctgaacaggggccccggccctggagccac 1126
          |  |||| || |||||  |||  ||  |  | |  ||  |  ||||  |||  |  |||
Db    227 caatggctataggaccgaatacggtcttaacgggtcaaagaaccccaaccccaggaccat 286

Qy   1127 cgcataacttgccggtgccccaattaacctcgctgcggcctgacataacacagccgccta 1186
          |   | |  |  |                              || |  ||   |   |
Db    287 cctctgatataac-----------------------ttctaaattagaccccactgagt 322

Qy   1187 gcaacagtaccactggattgattcctaccaacacgcctagaaactccccaggtgttcctg 1246
            ||||| || ||
Db    323 ctaacagcacgac---------------------------------------------- 335

Qy   1247 ttaagacaggacagagactcttcagtctcatccagggagctttccaagccatcaactcca 1306
           ||| || ||    | ||| || ||||||||||||||||||| |||||  | |||||||
Db    336 -taaaacggggacaaaacttttagtctcatccagggagcttttcaagctcttaactcca 394

Qy   1307 ccgaccctgatgccacttcttcttgttggctttgtctatcctcagggcctccttattatg 1366
          |    || || || || |||||||||||||||||  || | || || || || ||||| |||
Db    395 cgactccagaggctacctcttcttgttggctttgcttaacttcgggcccaccttactata 454

Qy   1367 aggggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacat 1426
          ||   |||||||||  |  |||||||||||||| ||| | |||||| | ||||||||||
Db    455 agaaaatggctaaaagagaaaaattcaatgtgacaaaaaaacatagagaccaatgtacat 514

Qy   1427 gggggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgcataggaa 1486
          |||| |||| |||||||||||||||| ||||| ||||| || ||||  || ||||||  ||
Db    515 ggggatcccaaaataagcttaccctactgaggtttctggaaaggacacctgcataaaaa 574

Qy   1487 aagctcccccatcccaccaacacctttgctatagtactgtggtttatgagcaggcctcag 1546
          | | ||||||||||||||||||||| | |    ||||  |  | | | ||  |||| |
Db    575 aggttcccccatcccaccaacacctttacaaccacactgaagcctttaatcaaacctctg 634

Qy   1547 aaaatcagtatttagtacctggttataacaggtggtgggcatgcaatactgggttaaccc 1606
          | | ||| ||| | |||||||||||| |||||||||||||||| |||||||| |||||||
Db    635 agagtcaatatctggtacctggttatgacaggtggtgggcatgtaatactggattaaccc 694

Qy   1607 cctgtgtttccacctcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcg 1666
          | |||||||||||  || |||||||| | ||||| || |    | ||||||||||| |
Db    695 cttgtgtttccaccttggttttcaaccaaactaaagacttttacattatggtccaaattg 754

Qy   1667 tcccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcggtata 1726
          | |||||||| || ||| ||||  || ||    | || |||||||||| ||  || |
```

```
Db    755 ttccccgagtatattactatcccaagaaaacaattctcgatgaatatgattacaggaacc 814

Qy   1727 accgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggattagggacgg 1786
          | |||| ||| | | ||||| ||||||| || || || |||||||||||| | ||   |
Db    815 atcgacaaaagaaaaaacccatatccctgacactcgcagtaatgctcggactcggagtga 874

Qy   1787 ccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaaag 1846
                || ||  ||||||| || || ||  |  | |||||||| |||||||||| | ||
Db    875 taacaggtgtgagaacaggaactgcagctttagttacaggacctcagcagctagaaacag 934

Qy   1847 gacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgtta 1906
          |||||  || | ||||||       | |||| | |||||| |||| ||||  | ||||| |
Db    935 gacttagtaacctacatcaaattgtaacaggaaatctccaagccctaaaaaaatctgtca 994

Qy   1907 gcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggaggggat 1966
          | ||||| ||| | ||||| || || || |||||||| ||||||||||    | || |
Db    995 gtaacctggaaaaatccctaacctccttatctgaagtagttctacagaataaaaaaggt 1054

Qy   1967 tagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgct 2026
          |||||  |  | ||||||| | || |  | |||||||  |||||||| ||  |||| || |
Db    1055 tagatttattatttctaaaaaaaagaagattatgtgtagccttaaaggagaaatgctgtt 1114

Qy   2027 tctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagttag 2086
          | |||||||||| ||||| |||||||||||||||||||||||||||||| ||||| ||| |
Db    1115 tttatgtagatcattcaggggccatcagagactccatgaacaagcttaaaaaaaggttgg 1174

Qy   2087 agaggcgtcgaagggaaagagaggctgac-caggggtggtttgaaggatggttcaacagg 2145
          |||  |||||||||||||  ||  || || ||  | |||||| || ||||||||||||||
Db    1175 agaaacgtcgaagggaaaaggaaacttactcaaagatggtttaaaagatggttcaacagg 1234

Qy   2146 tctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctcctg 2205
          ||||||||| || | ||||| |||||| || | || || ||| || ||||||| ||||||
Db    1235 tctccttggttggctaccctactttctactttaacaggacccttaatagtcctcctcctg 1294

Qy   2206 ttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacgagtg 2265
          ||||| |||||||||||  ||  | ||||| | |||  |||||||  ||||| ||||| |
Db    1295 ttactcacagttgggccatgtattattaacaagttaattgccttcattagaaaacgaata 1354

Qy   2266 agtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaaggagaa 2325
          |||||||||||||||||||||||||||| ||||||||||||  ||     |  | | |||
Db    1355 agtgcagtccagatcatggtacttagacaacagtaccaaaaccccatcta---acagggaa 1411

Qy   2326 actgacctctagccttcccagttctaagattagaactattaacaagacaagaagtgggga 2385
           ||| || |||||  | |||||||||||||| |||||||||||||| ||||||||||||
Db    1412 gctggccgctagctctaccagttctaagattaaaactattaacaagagaagaagtgggga 1471

Qy   2386 atgaaaggatga 2397
          |||||||||| |
Db    1472 atgaaaggataa 1483


RESULT   9
US-09-011-745-7
; Sequence 7, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 6312
;   TYPE: DNA
```

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4058)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4059)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4060)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4061)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4246)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4247)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4248)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4249)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-7


  Query Match             15.4%;  Score 455;  DB 4;  Length 6312;
  Best Local Similarity   53.0%;  Pred. No. 8.5e-137;
  Matches 1261;  Conservative    0;  Mismatches 1030;  Indels   88;  Gaps   10;

Qy       2 gctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgctcta 61
Db    1208 gctgcttagggccaggaataccctggccggtttggtttaactccttatgaaattctcta 1267

Qy      62 cgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgctttc 121
Db    1268 tggaggaccacccccatacttgagtctggagaaactttgggtcccgatgatagatttct 1327

Qy     122 ccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagca 181
Db    1328 ccctgtcttatttactcacttaaaggctttagaaattgtaaggacccaaatctgggacca 1387

Qy     182 gctccgggaggcctactcaggagagagagacttgcaagttccacatcgcttccaagttgg 241
Db    1388 gatcaaagaggtgtataagcctg---gtaccgtaacaatccctcacccgttccaggtcgg 1444

Qy     242 agattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggacc 301
Db    1445 ggatcaagtgcttgtcagacgccatcgacccagcagccttgagcctcggtggaaaggccc 1504

Qy     302 ttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctggat 361
Db    1505 atacctggtgttgctgactaccccgaccgcggtaaaagtcgatggtattgctgcctgggt 1564

Qy     362 ccatgcatcccacgttaagctggcgccacct------cccgactcggggtggagagccga 415
Db    1565 ccatgcttctcacctcaaacctgcaccaccttcggcaccagatgagtcctgggagctgga 1624

Qy     416 aaagactgagaatccccttaagcttcgcctccatcgcctggttccttactct-aacaata 474
Db    1625 aaagactgatcatcctcttaagctgcgtattcggcggcggcgggacgagtctgcaaaata 1684

Qy     475 actccccaggccagtagtaaacgccttatagacagctcgaaccccca----tagacctttt 530
Db    1685 agaaccccccaccagcccatgaccctcacttggcaggtactgtcccaaactggagacgttg 1744

Qy     531 atcccctacctggctgattattgaccctgatacgggtgtcactgtaaatagcactcgagg 590
Db    1745 tctgggatacaaaggcagtccagccccctggacttggtggcccacacttaaacctgatg 1804

Qy     591 tgttgctcctagaggcacctggtggcctgaactgcatttctgcctccgattgattaaccc 650
Db    1805 tatgtgccttggcggctagtcttgagtcctgggatatcccgggaaccgatgtctc--gtc 1862

Qy     651 cgctgttaaaagcacacctcccaacctagtccgtagttatgggttctattgctgcccagg 710
Db    1863 ctctaaacgagtcagacctccggactcagactatactgccgcttataagcaaatcacctg 1922

Qy     711 cacagagaaagagaaatactgtgggggttctggggaatccttctgtaggagatggagctg 770
Db    1923 gggagccatagggtgcagctaccctcgggctaggactagaatggcaagctctaccttcta 1982

Qy     771 cgtcacctccaacgatgg---------------------------------------- 788
Db    1983 cgtatgtccccgggatggccggacccttcagaagctagaaggtgcgggggctagaatc 2042

Qy     789 --------agactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtc 840
Db    2043 cctatactgtaaagaatgggattgtgagaccacggggaccggttattggctatctaaatc 2102

Qy     841 aattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctcccca 900
Db    2103 ctcaaaagacctcataactgtaaaatgggaccaaaatagcgaatggactcaaaaatttca 2162

Qy     901 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 960
Db    2163 acagtgtcaccagaccggctggtgtaacccccttaaaatagatttcacagacaaaggaaa 2222

Qy     961 a----------agtggataaatggtatgagctggggaatagttttttaataatatggcgg 1010
Db    2223 attatccaaggactggataacgggaaaaacctggggattaagattctatgtgtctggaca 2282

Qy    1011 gggagcagggtccactttaaccattcgccttaggatagagacggggacagaaccccctgt 1070
Db    2283 tc---caggcgtacagttcaccattcgcttaaaaatcaccaacatgccagctgtggcagt 2339

Qy    1071 ggcaatgggacccgataaagtactggctgaacaggggcccccggccctggagccaccgca 1130
Db    2340 aggtcctgacctcgtccttgtggaacaaggacctcctagaacgtccctcgctctcccacc 2399

Qy    1131 taacttgccggtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaa 1190
Db    2400 tcctcttcccccaagggaagcgccaccgccatctctccccgactctaactccacagccct 2459

Qy    1191 cagtaccactggattgattcctac------caacacgcctagaaactccccaggtgttcc 1244
Db    2460 ggcgactagtgcacaaactcccacggtgagaaaaacaattgttaccctaaacactccgcc 2519

Qy    1245 tgttaagacaggacagagactcttcagtctcatccagggagctttccaagccatcaactc 1304
Db    2520 tcccaccacaggcgacagacttttgatcttgtgcaggggcgccttcctaaccttaaatgc 2579

Qy    1305 caccgaccctgatgccacttcttcttgttggctttgtctatcctcagggcctccttatta 1364
Db    2580 taccaacccagggggccactgagtcttgctggctttgtttggccatgggccccccttatta 2639

Qy    1365 tgaggggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtac 1424
Db    2640 tgaagcaatagcctcatcaggagaggtcgcctactccaccgaccttg---accggtgccg 2696

Qy    1425 atgggggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgcatagg 1484
Db    2697 ctgggggacccaaggaaagctcaccctcactgaggtctcaggacacgggttgtgcatagg 2756
```

```
Qy   1485 aaaagctccccccatcccaccaacacctttgctatagtactgtggtttatgagcaggcctc 1544
          ||| |  |||    |||| || || || ||| ||   ||  |        |     ||
Db   2757 aaaggtgccctttacccatcagcatctctgcaatcagaccctatccatcaattcctccgg 2816

Qy   1545 agaaaatcagtatttagtacctggttataacaggtggtgggcatgcaatactgggttaac 1604
          |||  |||||||| |  | ||     |  | || |||||||| ||||  |||||  | ||
Db   2817 agaccatcagtatctgctcccctccaaccatagctggtgggcttgcagcactggcctcac 2876

Qy   1605 cccctgtgtttccacctcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaat 1664
          ||| ||  | ||||||||||| || || ||  | | |||||| ||| ||  ||||||  |
Db   2877 cccttgcctctccacctcagtttttaatcagactagagatttctgtatccaggtccagct 2936

Qy   1665 cgtcccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcggta 1724
            | || ||  | || ||| ||||||| |||||  |  |  | |  |||||| ||    |
Db   2937 gattcctcgcatctattactatcctgaagaagttttgttacaggcctatgacaattctca 2996

Qy   1725 taaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggattagggac 1784
            | |  | ||||||||  | || || |||||||||||||  | || || || || | |
Db   2997 ccccaggactaaaagagaggctgtctcacttaccctagctgttttactggggttgggaat 3056

Qy   1785 ggccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaa 1844
            | |  ||  |||| || ||  || ||||| | || | ||||||     | ||  || |
Db   3057 cacggcgggaataggtactggttcaactgccttaattaaaggacctatagacctccagca 3116

Qy   1845 aggacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgt 1904
          ||| ||       || ||    |||||    |  || ||||| ||| |  | || || ||
Db   3117 aggcctgacaagcctccagatcgccatagatgctgacctccgggccctccaagactcagt 3176

Qy   1905 tagcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggagggg 1964
           |||||  |||| || || ||||||||  |||| || || || || || ||  |||| ||
Db   3177 cagcaagttagaggactcactgacttccctgtccgaggtagtgctccaaaatagggagagg 3236

Qy   1965 attagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttg 2024
            | ||  |||||||||||| |||||||||  | ||||| ||| |||| ||||| || ||
Db   3237 ccttgacttgctgtttctaaaagaaggtggcctctgtgcggccctaaaggaagagtgctg 3296

Qy   2025 cttctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagtt 2084
           || ||  |||| |||||||| ||  |  | ||||||||||| || || | | ||||  |
Db   3297 tttttacatagaccactcaggtgcagtacgggactccatgaaaaaactcaaagaaaaact 3356

Qy   2085 agagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacag 2144
           || |   | |     ||  |  ||     |||    |||| |||||||||||||| |
Db   3357 ggataaaagacagttagagcgccagaaaagccaaaactggtatgaaggatggttcaataa 3416

Qy   2145 gtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctcct 2204
           || |||||| | || |||||||| ||  |  |  | ||||||||| || |||| || ||
Db   3417 ctccccttggttcactaccctgctatcaaccatcgctgggcccctattactcctccttct 3476

Qy   2205 gttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacgagt 2264
          ||| || |   | ||||| ||| | || ||| | || |||   ||||||| |||  |  |
Db   3477 gttgctcatcctcgggccatgcatcatcaatcgattagttcaatttgttaaagacaggat 3536

Qy   2265 gagtgcagtccagatcatggtacttaggcaacagtacca 2303
              |||||||    | || || |  ||||| |||||
Db   3537 ctcagtagtccaggctttagtcctgactcaacaatacca 3575


RESULT  10
US-08-716-351A-1
; Sequence 1, Application US/08716351A
; Patent No. 6033905
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Gibbon Ape Leukemia Virus-Based
;    TITLE OF INVENTION:  Retroviral Vectors
;    NUMBER OF SEQUENCES:  5
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/716,351A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/03784
;      FILING DATE:  06-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian,, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  15280-128-1PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..8535
;      OTHER INFORMATION:  /standard_name= "GaLV SEATO Genome"
US-08-716-351A-1


  Query Match             15.3%;  Score 451.2;  DB 3;  Length 8535;
  Best Local Similarity   52.8%;  Pred. No. 1.8e-135;
  Matches 1297;  Conservative    0;  Mismatches 1063;  Indels   98;  Gaps   11;

Qy       2 gctttttagggttaggaacacccctggacagtttgggctgacccccctatgaattgctcta 61
           |||  ||||||  ||||| |||||||| | ||||||  | || || |||||| | |||||
Db    5584 GCTGCTTAGGGCCAGGAATACCCCTGGCCGGTTTGGTTTAACTCCTTATGAAATTCTCTA 5643

Qy      62 cgggggaccccccccgttggtagaaattgcttctgtacatagtgctgatgtgctgctttc 121
            || ||||| |||||  |  | ||   ||            || | ||||      ||
Db    5644 TGGAGGACCACCCCCCATACTTGAGTCTGGAGAAACTTTGGGTCCCGATGATAGATTTCT 5703

Qy     122 ccagcctttgttctctaggctcaaggcgctcgagtgggtgaggcaacgagcgtggaagca 181
           ||      || ||  ||    | |||||  | ||    || |||   | |  ||| | ||
Db    5704 CCCTGTCTTATTTACTCACTTAAAGGCTTTAGAAATTGTAAGGACCCAAATCTGGGACCA 5763

Qy     182 gctccgggaggcctactcaggagagagagacttgcaagttccacatcgcttccaagttgg 241
           | ||   ||||  ||        |   |   | |   | | || || |  ||||| || ||
Db    5764 GATCAAAGAGGTGTATAAGCCTG---GTACCGTAACAATCCCTCACCCGTTCCAGGTCGG 5820

Qy     242 agattcagtctatgttagacgccaccgtgcaggaaacctcgagactcggtggaagggacc 301
           |||  |||   ||| |||||||| ||  |  | | ||| ||| |||||||||| || ||
Db    5821 GGATCAAGTGCTTGTCAGACGCCATCGACCCAGCAGCCTTGAGCCTCGGTGGAAAGGCCC 5880

Qy     302 ttatctcgtacttttgaccacaccaacggctgtgaaagtcgaaggaatctccacctggat 361
           || || ||  |  |||| || || || || || |||||||| || ||  |  ||||| |
Db    5881 ATACCTGGTGTTGCTGACTACCCCGACCGCGGTAAAAGTCGATGGTATTGCTGCCTGGGT 5940

Qy     362 ccatgcatcccacgttaagctggcgccacct------cccgactcggggtggagagccga 415
           |||||| || ||| | || |  || ||||||      || ||   |  |||      ||
Db    5941 CCATGCTTCTCACCTCAAACCTGCACCACCTTCGGCACCAGATGAGTCCTGGGAGCTGGA 6000

Qy     416 aaagactgagaatccccttaagcttcgcctccatcgcctggttccttactct-aacaata 474
           |||||||||  |||| |||||||| ||  | |  || | |       | |||  | ||||
Db    6001 AAAGACTGATCATCCTCTTAAGCTGCGTATTCGGCGGCGGCGGGACGAGTCTGCAAAATA 6060

Qy     475 actccccaggccagtagtaaacgccttatagacagctcgaaccccca----tagaccttt 530
           |   |||   ||||      || |    | | ||| |     ||| |     |||| ||
Db    6061 AGAACCCCCACCAGCCCATGACCCTCACTTGGCAGGTACTGTCCCAAACTGGAGACGTTG 6120

Qy     531 atcccctacctggctgattattgaccctgatacgggtgtcactgtaaatagcactcgagg 590
                     |      | |   | |||          ||  |   | |   ||  || | |
Db    6121 TCTGGGATACAAAGGCAGTCCAGCCCCCTTGGACTTGGTGGCCCACACTTAAACCTGATG 6180

Qy     591 tgttgctcctagaggcacctggtggcctgaactgcatttctgcctccgattgattaaccc 650
```

```
Db   6181 TATGTGCCTTGGCGGCTAGTCTTGAGTCCTGGGATATCCCGGGAACCGATGTCTCGTCCT 6240

Qy    651 cgctgttaaaagcacacctcccaacctagtccgtagttatgggttctattgctgcccagg 710
Db   6241 --CTAAACGAGTCAGACCTCCGGACTCAGACTATACTGCCGCTTATAAGCAAATCACCTG 6298

Qy    711 cacagagaaagagaaatactgtgggggttctggggaatccttctgtaggagatggagctg 770
Db   6299 GGGAGCCATAGGGTGCAGCTACCCTCGGGCTAGGACTAGAATGGCAAGCTCTACCTTCTA 6358

Qy    771 cgtcacctccaacgatgg---------------------------------------- 788
Db   6359 CGTATGTCCCCGGGATGGCCGGACCCTTTCAGAAGCTAGAAGGTGCGGGGGGCTAGAATC 6418

Qy    789 --------agactggaaatggccgatctctctccaggaccgggtaaaattctcctttgtc 840
Db   6419 CCTATACTGTAAAGAATGGGATTGTGAGACCACGGGGACCGGTTATTGGCTATCTAAATC 6478

Qy    841 aattccggcccgggcaagtacaaaatgatgaaactatataaagataagagctgctcccca 900
Db   6479 CTCAAAAGACCTCATAACTGTAAAATGGGACCAAAATAGCGAATGGACTCAAAAATTTCA 6538

Qy    901 tcagacttagattatctaaagataagtttcactgaaaaaggaaaacaggaaaatattcaa 960
Db   6539 ACAGTGTCACCAGACCGGCTGGTGTAACCCCCTTAAAATAGATTTCACAGACAAAGGAAA 6598

Qy    961 a----------agtggataaatggtatgagctggggaatagttttttaataatatggcgg 1010
Db   6599 ATTATCCAAGGACTGGATAACGGGAAAAACCTGGGGATTAAGATTCTATGTGTCTGGACA 6658

Qy   1011 gggagcagggtccactttaaccattcgccttaggatagagacggggacagaaccccctgt 1070
Db   6659 TC---CAGGCGTACAGTTCACCATTCGCTTAAAAATCACCAACATGCCAGCTGTGGCAGT 6715

Qy   1071 ggcaatgggacccgataaagtactggctgaacaggggccccggccctggagccaccgca 1130
Db   6716 AGGTCCTGACCTCGTCCTTGTGGAACAAGGACCTCCTAGAACGTCCCTCGCTCTCCCACC 6775

Qy   1131 taacttgccggtgccccaattaacctcgctgcggcctgacataacacagccgcctagcaa 1190
Db   6776 TCCTCTTCCCCCAAGGGAAGCGCCACCGCCATCTCTCCCCGACTCTAACTCCACAGCCCT 6835

Qy   1191 cagtaccactggattgattcctacc------aacacgcctagaaactccccaggtgttcc 1244
Db   6836 GGCGACTAGTGCACAAACTCCCACGGTGAGAAAAACAATTGTTACCCTAAACACTCCGCC 6895

Qy   1245 tgttaagacaggacagagactcttcagtctcatccagggagctttccaagccatcaactc 1304
Db   6896 TCCCACCACAGGCGACAGACTTTTTGATCTTGTGCAGGGGGCCTTCCTAACCTTAAATGC 6955

Qy   1305 caccgaccctgatgccacttcttcttgttggctttgtctatcctcagggcctccttatta 1364
Db   6956 TACCAACCCAGGGGCCACTGAGTCTTGCTGGCTTTGTTTGGCCATGGGCCCCCCTTATTA 7015

Qy   1365 tgaggggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtac 1424
Db   7016 TGAAGCAATAGCCTCATCAGGAGAGGTCGCCTACTCCACCGACCTTG---ACCGGTGCCG 7072

Qy   1425 atgggggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgcatagg 1484
Db   7073 CTGGGGGACCCAAGGAAAGCTCACCCTCACTGAGGTCTCAGGACACGGGTTGTGCATAGG 7132

Qy   1485 aaaagctcccccatcccaccaacacctttgctatagtactgtggtttatgagcaggcctc 1544
Db   7133 AAAGGTGCCCTTTACCCATCAGCATCTCTGCAATCAGACCCTATCCATCAATTCCTCCGG 7192

Qy   1545 agaaaatcagtatttagtacctggttataacaggtggtgggcatgcaatactgggttaac 1604
Db   7193 AGACCATCAGTATCTGCTCCCCTCCAACCATAGCTGGTGGGCTTGCAGCACTGGCCTCAC 7252

Qy   1605 cccctgtgtttccacctcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaat 1664
Db   7253 CCCTTGCCTCTCCACCTCAGTTTTTAATCAGACTAGAGATTTCTGTATCCAGGTCCAGCT 7312

Qy   1665 cgtcccccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcggta 1724
Db   7313 GATTCCTCGCATCTATTACTATCCTGAAGAAGTTTTGTTACAGGCCTATGACAATTCTCA 7372

Qy   1725 taaccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggattagggac 1784
Db   7373 CCCCAGGACTAAAAGAGAGGCTGTCTCACTTACCCTAGCTGTTTTACTGGGGTTGGGAAT 7432

Qy   1785 ggccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaa 1844
Db   7433 CACGGCGGGAATAGGTACTGGTTCAACTGCCTTAATTAAAGGACCTATAGACCTCCAGCA 7492

Qy   1845 aggacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgt 1904
Db   7493 AGGCCTGACAAGCCTCCAGATCGCCATAGATGCTGACCTCCGGGCCCTCCAAGACTCAGT 7552

Qy   1905 tagcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggagggg 1964
Db   7553 CAGCAAGTTAGAGGACTCACTGACTTCCCTGTCCGAGGTAGTGCTCCAAAATAGGAGAGG 7612

Qy   1965 attagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttg 2024
Db   7613 CCTTGACTTGCTGTTTCTAAAAGAAGGTGGCCTCTGTGCGGCCCTAAAGGAAGAGTGCTG 7672

Qy   2025 cttctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagtt 2084
Db   7673 TTTTTACATAGACCACTCAGGTGCAGTACGGGACTCCATGAAAAAACTCAAAGAAAAACT 7732

Qy   2085 agagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacag 2144
Db   7733 GGATAAAAGACAGTTAGAGCGCCAGAAAAGCCAAAACTGGTATGAAGGATGGTTCAATAA 7792

Qy   2145 gtctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctcct 2204
Db   7793 CTCCCCTTGGTTCACTACCCTGCTATCAACCATCGCTGGGCCCCTATTACTCCTCCTTCT 7852

Qy   2205 gttacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacgagt 2264
Db   7853 GTTGCTCATCCTCGGGCCATGCATCATCAATAAGTTAGTTCAATTCATCAATGATAGGAT 7912

Qy   2265 gagtgcagtccagatcatggtacttaggcaacagtaccaaggccttctgagccaaggaga 2324
Db   7913 AAGTGCATGTTAAAATTCTGGTCCTTAGACAAAATATCAGGCCCTAGAGAACGAAGGTAA 7972

Qy   2325 aactgacctctagccttcccagttctaagattagaactattaacaagacaagaagtgg 2382
Db   7973 CCTTTAATTTT----------GCTCTAAGATTAGAGCTATTCACAAGAGAAATGGGGG 8020
```

```
RESULT  11
US-09-075-272-1
; Sequence 1, Application US/09075272
; Patent No. 6136598
;  GENERAL INFORMATION:
;    APPLICANT:  MILLER, A. DUSTY
;    APPLICANT:  WOLGAMOT, GREG
;    APPLICANT:  BONHAM, LYNN
;    TITLE OF INVENTION:  MUS DUNNI ENDOGENOUS RETROVIRAL
;    TITLE OF INVENTION:  PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,272
;      FILING DATE:  08-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/046,140
;      FILING DATE:  09-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poor, Brian W.
;      REGISTRATION NUMBER:  32,928
;      REFERENCE/DOCKET NUMBER:  14538A-003710
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-075-272-1


  Query Match             12.6%;  Score 372.2;  DB 3;  Length 8655;
  Best Local Similarity   62.1%;  Pred. No. 8.3e-110;
  Matches  663;  Conservative    0;  Mismatches  383;  Indels   21;  Gaps    4;

Qy   1250 agacaggacagagactcttcagtctcatccagggagctttccaagccatcaactccaccg 1309
          | |||| | |  | ||  |||||||  | ||||||||  ||   ||   | ||    ||
Db   6706 ATACAGAAAACCGTCTGGTCAGTCTAGTTCAGGGAGCCTTTTTAGTTTTAAATAGGACTA 6765

Qy   1310 accctgatgccacttcttcttgttggctttgtctatcctcagggcctccttattatgagg 1369
          | ||| ||   |||   || |||||| | ||     ||||    || ||||||||||| |
Db   6766 ATCCTAATATGACTCAATCATGTTGGTTATGCTATGCCTCTAACCCCCCTTATTATGAAG 6825

Qy   1370 ggatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatggg 1429
          | || ||| |    ||||  | |||| | ||   ||| |||     |||||||   ||||
Db   6826 GAATTGCTCAGACTAGAACTTACAATATTACTTCAGATCAT---TCTCAATGTCTTTGGG 6882

Qy   1430 ggtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgcataggaaaag 1489
          |       | ||| | || || || | |||||| || || |||   ||  ||||  | |
Db   6883 GAGAGAACAGGAAGTTGACTCTGACAGCAGTTTCAGGAAATGGGCTTTGTTTAGGTCAGG 6942

Qy   1490 ctcccccatcccaccaacacctttgctatagtactgtggtttatgagcaggcctcagaaa 1549
            ||||      |    ||||| ||  |    |||     |       |     |  ||
Db   6943 TGCCCCAGGATAAATGGCACCTCTGTAACCAGACTCAAAATATCCGACCTAACAAAGGTG 7002

Qy   1550 atcagtatttagtacctggttataacaggtggtgggcatgcaatactgggttaacccccct 1609
           ||||||| |||| |||       |||     ||||| |||||||| ||  | || || |
Db   7003 GTCAGTATCTAGTGCCTCCCATAGACACAGTATGGGCTTGCAATACAGGTCTCACTCCTT 7062

Qy   1610 gtgtttccacctcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtcc 1669
          || | || |  || || |||||    |||||||||| ||| |  | || ||  |  | |
Db   7063 GTATATCTATGTCTGTTTTCAATAGCTCCAAAGATTTCTGTATTTTAGTTCAGCTTATTC 7122

Qy   1670 cccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcggtataacc 1729
          |  || |    || |||  |||       | || ||  ||| |||  |||||     ||
Db   7123 CTAGACTCCTGTATCATGATGATAGCTCATTTCTAGACAAATTTGAACATCGG---GTCC 7179

Qy   1730 gaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggatt---agggacgg 1786
          |    |||||||||||| |  |  | ||  |||| ||  |  | |||||   || | | |
Db   7180 GCTGGAAAAGAGAACCCATTACTTTAACGTTAGCAGTGCTTTTAGGATTGGGAGTGGCCG 7239

Qy   1787 ccgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaaag 1846
          | | ||| ||||| || || || |||||| | |||       || |   |  | |||
Db   7240 CAGCTGGGGTAGGTACTGGAACCGCTGCCTTAATCCAGACCCCCCGATACTTTGAG---- 7295

Qy   1847 gacttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgtta 1906
                ||  ||| | | || |||      |||||| |||  || |  | ||| | |
Db   7296 --------GAATTACGTACAGCTATGGATACTGATCTCAGAGCTATAGAACACTCTATAA 7347

Qy   1907 gcaacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggaggggat 1966
          ||| |||||||| ||  | |||||| ||||||| || || || || ||  |||| ||||
Db   7348 CCAAACTAGAAGAATCTTTAACTTCTCTGTCTGAGGTAGTACTGCAAAATAGGAGAGGAT 7407

Qy   1967 tagatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgct 2026
          ||||| || | |||||| | |||||| ||  | ||||| ||  | ||||| ||||||||
Db   7408 TAGATTTGTTATTTCTTAAAGAAGGAGGACTTTGTGCTGCTCTCAAAGAGGAATGTTGTT 7467

Qy   2027 tctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagttag 2086
          ||||||| || ||||| |||| |||| ||| || ||| ||| |||||| ||    |||
Db   7468 TCTATGTCGACCACTCGGGAGTGATCAAAGATTCTATGGCCAAACTTAGAGAACGCCTAG 7527

Qy   2087 agaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggt 2146
          | |  |||| ||| || ||||||    | || || |||||||| | ||||| || | ||
Db   7528 ATATACGTCAAAGAGAGAGAGAGCAAACAAGGATGGTTTGAAAGCTGGTTTAATAAGT 7587

Qy   2147 ctccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctcctgt 2206
          | || ||| | ||||| || |||||  || | ||||| ||| |  |  | || || || |
Db   7588 CCCCCTGGCTCACCACTCTCCTTTCCACTATAGCGGGACCCTTGATTATACTTCTGCTTT 7647

Qy   2207 tacttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacgagtga 2266
          | |||||  |||| || ||| |  |||||| ||| || |||||| ||||||| || | |
Db   7648 TGCTTACTTTTGGCCCCTGCATTCTTAATAAGTTAGTAGCCTTTATTAGAGAAAGAATAA 7707

Qy   2267 gtgcagtccagatcatggtacttaggcaacagtaccaaggccttctg 2313
          |||||| ||| | |||||| | |  ||||| || || | | | | |
Db   7708 ATGCAGTACAGGTTATGGTATTAAAACAACAATATCAGGTCTTCCAG 7754


RESULT  12
US-08-716-351A-5
; Sequence 5, Application US/08716351A
; Patent No. 6033905
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Gibbon Ape Leukemia Virus-Based
;    TITLE OF INVENTION:  Retroviral Vectors
;    NUMBER OF SEQUENCES:  5
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/716,351A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/03784
;      FILING DATE:  06-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian,, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  15280-128-1PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10970 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..10970
;      OTHER INFORMATION:  /standard_name= "p537 retroviral
;      OTHER INFORMATION:  vector"
US-08-716-351A-5


  Query Match             12.1%;  Score 357.8;  DB 3;  Length 10970;
```

```
  Best Local Similarity   62.9%;  Pred. No. 4.7e-105;
  Matches  592;  Conservative    0;  Mismatches  337;  Indels   12;  Gaps    2;

Qy   1471 gggacatgcataggaaaagctcccccatcccaccaacacctttgctatagtactgtggtt 1530
          ||| |||||||||| || |  || |   |||| ||||| |||||| |    ||  |
Db   6433 GGGCCATGCATAGGGAAGGTGCCTCTTACCCATCAACATCTTTGCAACCAGACCTTACCC 6492

Qy   1531 tatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggcatgc 1590
             |     |   | |  |||||||| |  | ||     |  | || |||||||| |||
Db   6493 ATCAATTCCTCTAAAAACCATCAGTATCTGCTCCCCTCAAACCATAGCTGGTGGGCCTGC 6552

Qy   1591 aatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttgtgt 1650
          |  |||||  | ||||||||  | ||||||||||| || || || || ||||| || |||
Db   6553 AGCACTGGCCTCACCCCCTGCCTCTCCACCTCAGTTTTTAATCAGTCTAAAGACTTCTGT 6612

Qy   1651 gtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgatgaa 1710
          |||  ||||||  |  ||||||| | || |||||| |||| |||    |  |  | |
Db   6613 GTCCAGGTCCAGCTGATCCCCCGCATCTATTACCATTCTGAAGAAACCTTGTTACAAGCC 6672

Qy   1711 tatgactatcggtataaccgaccaaaaagagaacccgtatcccttaccctagctgtaatg 1770
          ||||||  |          | |    |||||||| || |  || ||||||||||||||  |
Db   6673 TATGACAAATCACCCCCCAGGTTTAAAAGAGAGCCTGCCTCACTTACCCTAGCTGTCTTC 6732

Qy   1771 ctcggattagggacggccgttggcgtaggaacagggacagctgccctgatcacaggacca 1830
          || || |||||||  || |  ||  |||| || ||  || | ||||| || | ||| ||
Db   6733 CTGGGGTTAGGGATTGCGGCAGGTATAGGTACTGGCTCAACCGCCCTAATTAAAGGGCCC 6792

Qy   1831 cagcagctagagaaaggacttggtgagctacatgcggccatgacagaagatctccgagcc 1890
             | ||  || |||| ||        || ||    |||||     |  || ||||| |||
Db   6793 ATAGACCTTCAGCAAGGCCTAACCAGCCTCCAAATCGCCATTGACGCTGACCTCCGGGCC 6852

Qy   1891 ttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaagtggttcta 1950
            |  |||| ||  | ||||| ||||| || || |||||||  | ||||| || || ||
Db   6853 CTTCAGGACTCAATCAGCAAGCTAGAGGACTCACTGACTTCCCTATCTGAGGTAGTACTC 6912

Qy   1951 cagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtgcagcctta 2010
          || ||  |||| ||  | ||  | || || || | |||||| ||  | || || ||| ||
Db   6913 CAAAATAGGAGAGGCCTTGACTTACTATTCCTTAAAGAAGGAGGCCTCTGCGCGGCCCTA 6972

Qy   2011 aaagaagaatgttgcttctatgtagatcactcaggagccatcagagactccatgaacaag 2070
          |||||||| || || || |||||||| |||||||| ||  |  |||||||||||| ||
Db   6973 AAAGAAGAGTGCTGTTTTTATGTAGACCACTCAGGTGCAGTACGAGACTCCATGAAAAAA 7032

Qy   2071 cttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaa 2130
          |||| | |||   |||| |   | |     ||  |  ||   ||||   |||| ||||
Db   7033 CTTAAAGAAAGACTAGATAAAAGACAGTTAGAGCGCCAGAAAAACCAAAACTGGTATGAA 7092

Qy   2131 ggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggccccta 2190
          || |||||||| |  || |||||| | || ||||| || ||  |  |  | |||||||||
Db   7093 GGGTGGTTCAATAACTCCCCTTGGTTTACTACCCTACTATCAACCATCGCTGGGCCCCTA 7152

Qy   2191 gtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgccttt 2250
           |  |||| ||  |||||||  ||  ||||||| ||| | || |||| ||  |    ||
Db   7153 TTGCTCCTCCTTTTGTTACTCACTCTTGGGCCCTGCATCATCAATAAATTAATCCAATTC 7212

Qy   2251 gttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggcctt 2310
          | |  ||  |  | |||||||||| | ||  | || ||||| ||  | ||  ||   |||
Db   7213 ATCAATGATAGGATAAGTGCAGTCAAAATTTTAGTCCTTAGACAGAAATATCAGACCCTA 7272

Qy   2311 ctgagccaaggagaaactgacctctagccttcccagttctaagattagaactattaacaa 2370
          | | | | | |   | |  |           | ||||||||||| ||||  ||||
Db   7273 GATAACGAGGAAAACCTTTAATT----------TCGCTCTAAGATTAGAGCTATCCACAA 7322

Qy   2371 gacaagaagtggggaatgaaaggatgaaaatgcaacctaac 2411
          || |  ||  |||||||||||| |      | |||  || |
Db   7323 GAGA--AATGGGGGAATGAAAGAAGTGTTTTTCAAGTTAGC 7361


RESULT  13
US-09-376-781-18
; Sequence 18, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 698
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Sequence of a
; Patent No. 6261806
;   OTHER INFORMATION: 5' env fragment of PERV-D.
US-09-376-781-18


  Query Match             11.6%;  Score 343;  DB 4;  Length 698;
  Best Local Similarity   71.0%;  Pred. No. 4.1e-101;
  Matches  483;  Conservative    0;  Mismatches  191;  Indels    6;  Gaps    2;

Qy    322 caccaacggctgtgaaagtcgaaggaatctccacctggatccatgcatcccacgttaagc 381
          ||||||||||||||||||||||||  ||||||||||| |||||| ||||||||||||||
Db      1 caccaacggctgtgaaagtcgaaaaaatctccacctagatccatacatcccacgttaagc 60

Qy    382 tggcgccacctcccgactcggggtggagagccgaaaagactgagaatccccttaagcttc 441
          ||||||||||||| || || ||||| | |||| |||||||| | ||||||||||||||||
Db     61 cggcgccacctcctgattcagggtgaaaagccaaaaagactaaaaatccccttaagcttc 120

Qy    442 gcctccatcgcctggttccttactctaacaataactccccaggccagtagtaaacgcctt 501
          |||||||||  | ||||||||||||||  |||||||  | ||| | | | |||  | | |
Db    121 gcctccatcacatggttccttactctgtcaataacctctcagactaatggtatgcacata 180

Qy    502 atagacagctcgaaccccccatagacctttatcccctacctggctgattattgaccctgat 561
            |||||||  |||| |||||| ||| |||||  |  |||||| | |||| |||| ||||
Db    181 ggagacagcctgaactcccataaaccctatctctgacctggttaattactgactctgac 240

Qy    562 acgggtgtcactgtaaatagcactcgaggtgttgctcctagaggcacctggtggcctgaa 621
          || ||| | | | |  | ||| ||||||| |  ||||||  ||  ||||||||||||||
Db    241 acaggtattaatatncacagcgctcgaggggaggctcctttagaaacctggtggcctgat 300

Qy    622 ctgcatttctgcctccgattgattaaccccgctgttaaaagcacacctcccaacctagtc 681
          ||  || |||||||| |||   | |  |   ||  |   | | || | ||  |  |  ||
Db    301 ctatatgtctgcctcagatcagtcattc---ctagtctgacctcaaccccagatatcctc 357

Qy    682 cgtagttatgggttctattgctgcccagg---cacagagaaagagaaatactgtgggggt 738
          |||  ||| || || |||   ||||||||     || | || |  ||| ||| |||   |
Db    358 cgtgcttacggattttatgtttgcccaggaccaccaaataatggaaaacactatggaaat 417

Qy    739 tctggggaatccttctgtaggagatggagctgcgtcacctccaacgatggagactggaaa 798
          || | || | ||| |  |    ||||||||| || ||||| || |||||| |  | |||
Db    418 cctagagatttcttttacaaacaatggagctgtgtaacctctaatgatggaaatcgaaaa 477

Qy    799 tggccgatctctctccaggaccgggtaaaattctcctttgtcaattccggcccgggcaag 858
          |||||  | ||||||  |||||  ||||||  || || | ||||||  ||        |  |
Db    478 tggccaacctctctgcaggatagggtaagcttttcttatgtcaaccccaataaccaccgg 537

Qy    859 tacaaaatgatgaaactatataaagataagagctgctccccatcagacttagattatcta 918
            |   |  | | |       |  ||    | || |  || |||||| ||||||||||
Db    538 acctggaaaacgtcatacaggtaccattctgggtgttttccctcagacctagattatctt 597

Qy    919 aagataagtttcactgaaaaaggaaaacaggaaaatattcaaaagtggataaatggtatg 978
          || |||||||||||  ||||||  ||||||  |||||||| | ||| ||||||||||||||
Db    598 aaaataagtttcaccgaaaaaaaaaaacaagaaaatatcctaaaatggataaatggtatg 657

Qy    979 agctggggaatagtttttta 998
```

```
Db    658 tcctggggaataatatatta 677


RESULT  14
US-09-011-745-5
; Sequence 5, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 6028
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3774)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3775)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3776)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3777)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3962)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3963)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3964)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3965)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-5


  Query Match             10.4%;  Score 307;  DB 4;  Length 6028;
  Best Local Similarity   56.7%;  Pred. No. 9.5e-89;
  Matches  631;  Conservative    0;  Mismatches  470;  Indels   12;  Gaps    3;

Qy   1191 cagtaccactggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaa 1250
Db   2191 cactaccagtacaccctcaacctcccctacaagtccaagtgtcccacagccacccccagg 2250

Qy   1251 gacaggacagagactcttcagtctcatccagggagctttccaagccatcaactccaccga 1310
Db   2251 aactggagatagactactagctctagtcaaaggagcctatcaggcgcttaacctcaccaa 2310

Qy   1311 ccctgatgccacttcttcttgttggctttgtctatcctcagggcctccttattatgaggg 1370
Db   2311 tcccgacaagacccaagaatgttggctgtgcttagtgtcgggacctccttattacgaagg 2370

Qy   1371 gatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggg 1430
Db   2371 agtagcggtcgtgggcacttataccaatcattccaccgctccggccaactgtacggccac 2430

Qy   1431 gtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagc 1490
Db   2431 ttcccaacataagcttaccctatctgaagtgacaggacagggcctatgcatgggggcagt 2490

Qy   1491 tcccccatcccaccaacaccttgctatagtactgtggtttatgagcaggcctcagaaaa 1550
Db   2491 acctaaaactcaccaggccttatgtaacaccac------ccaaagcgccggctcaggatc 2544

Qy   1551 tcagtatttagtacctggttataacaggtggtgggcatgcaatactgggttaacccccctg 1610
Db   2545 ctactaccttgcagcacccgccggaacaatgtgggcttgcagcactggattgactccctg 2604

Qy   1611 tgtttccacctcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtccc 1670
Db   2605 cttgtccaccacggtgctcaatctaaccacagattattgtgtattagttgaactctggcc 2664

Qy   1671 ccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcggtataaccg 1730
Db   2665 cagagtaatttaccactcccccgattatatgtatggtcagcttgaacagcg---taccaa 2721

Qy   1731 accaaaaagagaacccgtatcccttaccctagctgtaatgctcggattagggacggccgt 1790
Db   2722 atataaaagagagccagtatcattgaccctggcccttctactaggaggattaaccatggg 2781

Qy   1791 tggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaaaggact 1850
Db   2782 agggattgcagctggaatagggacggggaccactgccttaattaaaacccagca---gtt 2838

Qy   1851 tggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgttagcaa 1910
Db   2839 tgagcagcttcatgccgctatccagacagacctcaacgaagtcgaaaagtcaattaccaa 2898

Qy   1911 cctagaagagtccctgacttctttgtctgaagtggttctacagaaccggaggggattaga 1970
Db   2899 cctagaaaagtcactgacctcgttgtctgaagtagtcctacagaaccgcagaggcctaga 2958

Qy   1971 tctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttcta 2030
Db   2959 tttgctattcctaaaggagggaggtctctgcgcagccctaaaagaagaatgttgttttta 3018

Qy   2031 tgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagttagagag 2090
Db   3019 tgcagaccacacggggctagtgagagacagcatggccaaattaagagaaaggcttaatca 3078

Qy   2091 gcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctcc 2150
Db   3079 gagacaaaaactatttgagacaggccaaggatggttcgaagggctgtttaatagatcccc 3138

Qy   2151 ttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctcctgttact 2210
Db   3139 ctggtttaccaccttaatctccaccatcatgggacctctaatagtactcttactgatctt 3198

Qy   2211 tacagttgggccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgc 2270
Db   3199 actctttggaccttgcattctcaatcgattagttcaatttgttaaagacaggatctcagt 3258

Qy   2271 agtccagatcatggtacttaggcaacagtacca 2303
Db   3259 agtccaggctttagtcctgactcaacaatacca 3291
```

```
RESULT  15
US-08-258-420-9
; Sequence 9, Application US/08258420
; Patent No. 5710037
;  GENERAL INFORMATION:
;    APPLICANT:  Nienhuis, Arthur W.
;    APPLICANT:  Vanin, Elio F.
;    TITLE OF INVENTION:  No. 5710037el Retroviral Envelope and LTR and Retroviral Vecto
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan,Cecchi, Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch diskette
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,420
;      FILING DATE:  10-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME: -Olstein, Elliot M.
;      REGISTRATION NUMBER:  24,025
;      REFERENCE/DOCKET NUMBER:  271010-208
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1965 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-258-420-9
```

Query Match 10.3%; Score 303.8; DB 1; Length 1965;
Best Local Similarity 56.5%; Pred. No. 4.6e-88;
Matches 629; Conservative 0; Mismatches 472; Indels 12; Gaps 3;

```
Qy   1191 cagtaccactggattgattcctaccaacacgcctagaaactccccaggtgttcctgttaa 1250
          || ||||| |  |       |  ||   ||   |  ||    ||||    | ||
Db    834 CACTACCAGTACACCCTCAACCTCCCCTACAAGTCCAAGTGTCCCACAGCTACCCCCAGG 893

Qy   1251 gacaggacagagactcttcagtctcatccagggagctttccaagccatcaactccaccga 1310
          || ||| | |||||  |   |||  || | ||||| |  || ||  | |||  |||| |
Db    894 AACTGGAGATAGACTACTAGCTCTAGTCAAAGGAGCCTATCAGGCGCTTAACCTCACCAA 953

Qy   1311 ccctgatgccacttcttcttgttggctttgtctatcctcagggcctccttattatgaggg 1370
           || ||    ||       |||||||| ||  ||   || || ||||||||||| || ||
Db    954 TCCCGACAAGACCCAAGAATGTTGGCTGTGCTTAGTGTCGGGACCTCCTTATTACGAAGG 1013

Qy   1371 gatggctaaagaaagaaaattcaatgtgaccaaagagcatagaaatcaatgtacatgggg 1430
          | ||    |  ·| |  |  |                |       | |||||
Db   1014 AGTAGCGGTCGTGGGCACTTATACCAATCATTCCACCGCTCCGGCCAACTGTACGGCCAC 1073

Qy   1431 gtcccgaaataagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagc 1490
          ||||  | ||||||||||||||  |||||||  | ||  ||||   |||||| |||  ||
Db   1074 TTCCCAACATAAGCTTACCCTATCTGAAGTGACAGGACAGGGCCTATGCATGGGAGCAGT 1133

Qy   1491 tcccccatcccaccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaa 1550
           ||   | | |||||   | | ||  | |  ||        |       | ||||| |
Db   1134 ACCTAAAACTCACCAGGCCTTATGTAACACCAC------CCAAAGCGCCGGCTCAGGATC 1187

Qy   1551 tcagtatttagtacctggttataacaggtggtgggcatgcaatactgggttaacccccctg 1610
            | ||  | | | |          |    |||||| ||||  ||||| || || |||||
Db   1188 CTACTACCTTGCAGCACCCGCCGGAACAATGTGGGCTTGCAGCACTGGATTGACTCCCTG 1247

Qy   1611 tgtttccacctcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtccc 1670
           | |||||| | ||  |||| | | ||| |||||  |||||  | ||  || ||  ||
Db   1248 CTTGTCCACCACGGTGCTCAATCTAACCACAGATTATTGTGTATTAGTTGAACTCTGGCC 1307

Qy   1671 ccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcggtataaccg 1730
           | ||||    |||||  |    ||     ·|   || | |   |||  | ||   || |
Db   1308 CAGAGTAATTTACCACTCCCCCGATTATATGTATGGTCAGCTTGAACAGCG---TACCAA 1364

Qy   1731 accaaaaagagaacccgtatcccttaccctagctgtaatgctcggattagggacggccgt 1790
           |   |||||||| || |||||  | |||||  ||  |  | ||  |||  |   ||   |
Db   1365 ATATAAAAGAGAGCCAGTATCATTGACCCTGGCCCTTCTACTAGGAGGATTAACCATGGG 1424

Qy   1791 tggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaaaggact 1850
           ||  | | | | || | ||    |   || ||| |    |   |   |   |||| |   |
Db   1425 AGGGATTGCAGCTGGAATAGGGACGGGGACCACTGCCTTAATTAAAACCCAGCA---GTT 1481

Qy   1851 tggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgttagcaa 1910
           ||   |||| ||||| || ||       ||| |||   |   |  |  ||||  ||| |||
Db   1482 TGAGCAGCTTCATGCCGCTATCCAGACAGACCTCAACGAAGTCGAAAAGTCAATTACCAA 1541

Qy   1911 cctagaagagtccctgacttctttgtctgaagtggttctacagaaccggaggggattaga 1970
           ||||||| |||| ||||| || |||||||||||  || ||||||||||| || ||  ||||
Db   1542 CCTAGAAAAGTCACTGACCTCGTTGTCTGAAGTAGTCCTACAGAACCGCAGAGGCCTAGA 1601

Qy   1971 tctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttcta 2030
           | |||| || ||||  || || ||  | || ||||||  |||||||||||||||||| || ||
Db   1602 TTTGCTATTCCTAAAGGAGGGAGGTCTCTGCGCAGCCCTAAAAGAAGAATGTTGTTTTTA 1661

Qy   2031 tgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagttagagag 2090
           || ||| ||| | ||      | ||||||  ||||  |||  | ||| ||| | |  |
Db   1662 TGCAGACCACACGGGGCTAGTGAGAGACAGCATGGCCAAATTAAGAGAAAGGCTTAATCA 1721

Qy   2091 gcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctcc 2150
           | | | ||    |   ||| | | ||| || ||||| |||||   ||| || || || ||
Db   1722 GAGACAAAAACTATTTGAGACAGGCCAAGGATGGTTCGAAGGGCTGTTTAATAGATCCCC 1781

Qy   2151 ttggatgaccaccctgcttttctgctctgacggggcccctagtagtcctgctcctgttact 2210
            ||| | ||||||  |  | ||  |  | | ||| || ||| |||| || | ||| |  |
Db   1782 CTGGTTTACCACCTTAATCTCCACCATCATGGGACCTCTAATAGTACTCTTACTGATCTT 1841

Qy   2211 tacagttgggccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgc 2270
           |||| |||||| |  | ||| | || ||    |||||| |||  |  |   |
Db   1842 ACTCTTTGGACCTTGCATTCTCAATCGATTAGTCCAATTTGTTAAAGACAGGATATCAGT 1901

Qy   2271 agtccagatcatggtacttaggcaacagtacca 2303
           ||||||    | ||  | |  |||| || ||
Db   1902 GGTCCAGGCTCTAGTTTTGACTCAACAATATCA 1934
```

Search completed: February 24, 2002, 03:01:22
Job time: 21433 sec

```
OM of: US-09-171-553B-10 to: Issued_Patents_NA:*   out_format : pfs

Date: Feb 24, 2002 10:19 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09171553/runat_22022002_122826_11171/app_query.fasta_1.3312
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=feb22std.rni
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09171553_@CGN1_1_158 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-171-553B-10
Query length: 657
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 134.760000



score_list:
Sequence          Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-3 + 3484.00 5811.28  4.8e-316  3482  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-1 + 2631.50 4383.03  1.7e-236  2462  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-2 + 2537.50 4211.27  6.4e-227  7333  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-1 + 2402.00 3982.36  3.6e-214  8060  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-3 + 2395.00 3970.49  1.7e-213  8132  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-25 + 2169.00 3608.43  2.4e-193  2000  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-30 + 2169.00 3608.43  2.4e-193  2000  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-24 + 1675.00 2781.95  2.6e-147  1493  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-7 + 1544.50 2544.42  4.5e-134  6312  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-1 + 1522.50 2503.64  8.4e-132  8535  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-075-272-1 + 1478.50 2429.52  1.1e-127  8655  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-8 + 1329.00 2178.78  1.0e-113  8323  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-8 + 1329.00 2178.78  1.0e-113  8323  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-8 + 1329.00 2178.78  1.0e-113  8323
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-6 + 1284.00 2107.17  1.0e-109  6061  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-850-961-1 + 1284.00 2103.14  1.7e-109  8332  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-5 + 1263.00 2071.94  9.3e-108  6028  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-7 + 1261.00 2083.08  2.2e-108  1914  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-850-961-13 + 1260.50 2081.68  2.7e-108  2001  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-9 + 1259.50 2080.23  3.2e-108  1965  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-9 + 1238.00 2035.35  1.0e-105  3925  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-13 - 1226.00 2005.87  4.4e-104  8202  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-258-420-8 + 1216.00 2007.48  3.6e-104  1911  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-552-369-19 + 1212.50 2001.16  8.1e-104  1979  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-105-483-310 + 1201.50 1979.72  1.3e-102  2499  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-709-209-310 + 1201.50 1979.72  1.3e-102  2499  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-101-310 + 1201.50 1979.72  1.3e-102  2499  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-5 + 1018.00 1652.65  2.1e-84  10970  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-18 +  671.00 1104.34  7.3e-54  698  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-278-630A-7 +  658.00 1073.63  3.8e-52  1407  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-484-126-2 +  658.00 1073.29  3.9e-52  1446  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-278-630A-9 +  634.50 1034.44  5.7e-50  1374  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-073-836-1 +  634.50 1017.84  4.8e-49  5109  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-073-836-2 +  634.50 1017.67  4.9e-49  5176  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-278-630A-8 +  629.00 1025.62  1.8e-49  1329  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-484-126-4 +  623.50 1016.12  6.0e-49  1356  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-8 +  480.00  795.09  1.2e-36  276  !  S
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-10 +  434.50  718.54  2.3e-32  278  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-4 +  433.00  670.35  1.1e-29  10306  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-1 +  428.00  679.76  3.3e-30  2518  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-8 +  428.00  669.07  1.3e-29  5865
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-007-282B-1 +  426.50  682.17  2.4e-30  1704
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-3 +  420.50  650.16  1.5e-28  9661
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-7 +  389.00  636.13  8.8e-28  445
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-078-311-20 +  297.00  468.81  1.8e-18  1216

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-3

seq_documentation_block:
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
;  APPLICANT: Stoye, Jonathan P
;  APPLICANT: Weiss, Robin A
;  TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
;  TITLE OF INVENTION: specific sequences
;  FILE REFERENCE: 4238/75168
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 3482
;   TYPE: DNA
;   ORGANISM: Porcine retrovirus
US-09-111-085-3


alignment_scores:
            Quality: 3484.00              Length:    657
              Ratio:  5.319                Gaps:      0
 Percent Similarity:  99.696   Percent Identity: 98.478

alignment_block:
 US-09-171-553B-10 x US-09-111-085-3   ..

 Align seg 1/1  to: US-09-111-085-3  from: 1  to: 3482

       1 MetHisProThrLeuSerTrpArgHisLeuProThrArgGlyGlyGluPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     911 ATGCATCCCACGTTAAGCTGGCGCCACCTCCCGACTCGGGGTGGAGAGCC 960

      17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     961 GAAAAGACTGAGAATCCCCTTAAGCTTCGCCTCCATCGCCTGGTTCCTTA 1010

      34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1011 CTCTAACAATAACTCCCCAGGCCAGTAGTAAACGCCTTATAGACAGCTCG 1060

      51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAspTh 67
         ||||||||||||||||||||    ||||||||||||||||||||||||||
    1061 AACCCCCATAGACCTTTATCCCTTACCTGGCTGATTATTGACCCTGATAC 1110

      67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1111 GGGTGTCACTGTAAATAGCACTCGAGGTGTTGCTCCTAGAGGCACCTGGT 1160

      84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1161 GGCCTGAACTGCATTTCTGCCTCCGATTGATTAACCCCGCTGTTAAAAGC 1210

     101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyTh 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1211 ACACCTCCCAACCTAGTCCGTAGTTATGGGTTCTATTGCTGCCCAGGCAC 1260

     117 rGluLysGluLysTyrCysGlyGlySerGlyGluSerPheCysArgArgT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1261 AGAGAAAGAGAAATACTGTGGGGGTTCTGGGGAATCCTTCTGTAGGAGAT 1310

     134 rpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSerLeu 150
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1311 GGAGCTGCGTCACCTCCAACGATGGAGACTGGAAATGGCCGATCTCTCTC 1360

 151 GlnAspArgValLysPheSerPheValAsnSerGlyProGlyLysTyrLy 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1361 CAGGACCGGGTAAAATTCTCCTTTGTCAATTCCGGCCCGGGCAAGTACAA 1410

 167 sMetMetLysLeuTyrLysAspLysSerCysSerProSerAspLeuAspT 184
     |:::||||||||||||||||||||||||||||||||||||||||||||||
1411 AGTGATGAAACTATATAAAGATAAGAGCTGCTCCCCATCAGACTTAGATT 1460

 184 yrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLys 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1461 ATCTAAAGATAAGTTTCACTGAAAAAGGAAAACAGGAAAATATTCAAAAG 1510

 201 TrpIleAsnGlyMetSerTrpGlyIleValPheTyrLysTyrGlyGlyGl 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1511 TGGATAAATGGTATGAGCTGGGGAATAGTTTTTTATAAATATGGCGGGGG 1560

 217 yAlaGlySerThrLeuThrIleArgLeuArgIleGluThrGlyThrGluP 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1561 AGCAGGGTCCACTTTAACCATTCGCCTTAGGATAGAGACGGGGACAGAAC 1610

 234 roProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyProPro 250
     ||||||||||:::||||||||||||||||||||||||||||||||||||||
1611 CCCCTGTGGCAGTGGGACCCGATAAAGTACTGGCTGAACAGGGGCCCCCG 1660

 251 AlaLeuGluProProHisAsnLeuProValProGlnLeuThrSerLeuAr 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1661 GCCCTGGAGCCACCGCATAACTTGCCGGTGCCCCAATTAACCTCGCTGCG 1710

 267 gProAspIleThrGlnProProSerAsnSerThrThrGlyLeuIleProT 284
     ||||||||||||||||||||||||||||:::||||||||||||||||||||
1711 GCCTGACATAACACAGCCGCCTAGCAACGGTACCACTGGATTGATTCCTA 1760

 284 hrAsnThrProArgAsnSerProGlyValProValLysThrGlyGlnArg 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1761 CCAACACGCCTAGAAACTCCCCAGGTGTTCCTGTTAAGACAGGACAGAGA 1810

 301 LeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSerThrAspPr 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1811 CTCTTCAGTCTCATCCAGGGAGCTTTCCAAGCCATCAACTCCACCGACCC 1860

 317 oAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProProTyrT 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1861 TGATGCCACTTCTTCTTGTTGGCTTTGTCTATCCTCAGGGCCTCCTTATT 1910

 334 yrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGluHisArg 350
     ||||||||||||||||||  ||||||||||||||||||||||||||||||
1911 ATGAGGGGATGGCTAAAGAAGGAAAATTCAATGTGACCAAAGAGCATAGA 1960

 351 AsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThrGluValSe 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1961 AATCAATGTACATGGGGGTCCCGAAATAAGCTTACCCTCACTGAAGTTTC 2010

 367 rGlyLysGlyThrCysIleGlyLysAlaProProSerHisGlnHisLeuC 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2011 CGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTT 2060

 384 ysTyrSerThrValValTyrGluGlnAlaSerGluAsnGlnTyrLeuVal 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2061 GCTATAGTACTGTGGTTTATGAGCAGGCCTCAGAAAATCAGTATTTAGTA 2110

 401 ProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLeuThrProCysVa 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2111 CCTGGTTATAACAGGTGGTGGGCATGCAATACTGGGTTAACCCCCTGTGT 2160

 417 lSerThrSerValPheAsnGlnSerLysAspLeuCysValMetValGlnI 434
     ||||||||||||||||||||||||||||||||:::||||||||||||||||
2161 TTCCACCTCAGTCTTCAACCAATCCAAAGATTTCTGTGTCATGGTCCAAA 2210

 434 leValProArgValTyrTyrHisProGluGluValValLeuAspGluTyr 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||

2211 TCGTCCCCCGAGTGTACTACCATCCTGAGGAAGTGGTCCTTGATGAATAT 2260

 451 AspTyrArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAl 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2261 GACTATCGGTATAACCGACCAAAAAGAGAACCCGTATCCCTTACCCTAGC 2310

 467 aValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGlyThrAlaA 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2311 TGTAATGCTCGGATTAGGGACGGCCGTTGGCGTAGGAACAGGGACAGCTG 2360

 484 laLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHis 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2361 CCCTGATCACAGGACCACAGCAGCTAGAGAAAGGACTTGGTGAGCTACAT 2410

 501 AlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAsnLe 517
     ||||||||||||||||||||||||||||||||:::||||||||||||||||
2411 GCGGCCATGACAGAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACCT 2460

 517 uGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgG 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2461 AGAAGAGTCCCTGACTTCTTTGTCTGAAGTGGTTCTACAGAACCGGAGGG 2510

 534 lyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLys 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2511 GATTAGATCTGCTGTTTCTAAGAGAAGGTGGGTTATGTGCAGCCTTAAAA 2560

 551 GluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMe 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2561 GAAGAATGTTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTCCAT 2610

 567 tAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArgGluAlaA 584
     |:::||||||||||::::::||||||||||||||||||||||||||||||
2611 GAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGAAAGAGAGGCTG 2660

 584 spGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMetThrThr 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2661 ACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACCACC 2710

 601 LeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuTh 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2711 CTGCTTTCTGCTCTGACGGGACCCCTAGTAGTCCTGCTCCTGTTACTTAC 2760

 617 rValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgV 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2761 AGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAACGAG 2810

 634 alSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeuLeu 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2811 TGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTTCTG 2860

 651 SerGlnGlyGluThrAspLeu 657
     |||||||||||||||||||||
2861 AGCCAAGGAGAAACTGACCTC 2881
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-1

```
seq_documentation_block:
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
;  APPLICANT: Stoye, Jonathan P
;  APPLICANT: Weiss, Robin A
;  TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
;  TITLE OF INVENTION: specific sequences
;  FILE REFERENCE: 4238/75168
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2462
```

```
;   TYPE: DNA
;   ORGANISM: Porcine retrovirus
US-09-111-085-1


alignment_scores:
            Quality: 2631.50              Length:    679
              Ratio:  4.468                Gaps:      9
 Percent Similarity:  86.745   Percent Identity: 73.049

alignment_block:
 US-09-171-553B-10 x US-09-111-085-1   ..

 Align seg 1/1  to: US-09-111-085-1  from: 1  to: 2462

       1 MetHisProThrLeuSerTrpArgHisLeuProThrArgGlyGlyGluPr 17
         |||||||||||||||||   ||||||||||   |||||||||:::||
     211 ATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCGGGGTGGAAAGCC 260

      17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
         |||||||||:::|||||||||||||||||||||||||||||||||||||
     261 GAAAAGACTGAAAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTA 310

      34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
         |||||:::|||||||||||:::::::::|||||||||:::||||||
     311 CTCTGTCAATAACTCCTCAAGTTAATGGTAAACGCCTTGTGGACAGCCCG 360

      51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAspTh 67
         |||   |||:::|||||||||   ||||||||::   |||      ||
     361 AACTCCCATAAACCCTTATCTCTCACCTGGTTACTTACTGACTCCGGTAC 410

      67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
         ||||:::::::::|||||||||:::|||   ||||||   ||||||||||
     411 AGGTATTAATATTAACAGCACTCAAGGGGAGGCTCCCTTGGGGACCTGGT 460

      84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaVal...... 98
         ||||||||||:::   |||||||||   :::   |||::::::
     461 GGCCTGAATTATATGTCTGCCTTCGATCAGTAATCCCTGGTCTCAATGAC 510

      99 LysSerThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysPr 115
         ::::::|||||||||:::::::::|||:::||||||||||||   |||||
     511 CAGGCCACACCCCCCGATGTACTCCGTGCTTACGGGTTTTACGTTTGCCC 560

     115 oGly...ThrGluLysGluLysTyrCysGlyGlySerGlyGluSerPheC 131
         ||||      ::::::|||:::|||||||||:::      :::   ||||
     561 AGGACCCCCAAATAATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTT 610

     131 ysArgArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpPro 147
         ||::::::|||||||||:::|||||||||||||||:::||||||||||||
     611 GCAAGCAATGGAGCTGCATAACTTCTAATGATGGGAATTGGAAATGGCCA 660

     148 IleSerLeuGlnAspArgValLysPheSerPheValAsnSer........ 161
         :::|||   ||||||||||||:::::|||||||||||:::
     661 GTCTCTCAGCAAGACAGAGTAAGTTACTCTTTTGTTAACAATCCTACCAG 710

     162 ...................GlyProGlyLysTyrLys......MetMetL 170
                           |||   |||:::::|||      :::::::
     711 TTATAATCAATTTAATTATGGCCATGGGAGATGGAAAGATTGGCAACAGC 760

     170 ysLeuTyrLysAspLys..............SerCysSerProSerAsp 181
         :::::   ||||||                |||||          |||
     761 GGGTACAAAAAGATGTACGAAATAAGCAAATAAGCTGTCATTCGTTAGAC 810

     182 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIl 198
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     811 CTAGATTACTTAAAAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATAT 860

     198 eGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTyrLysTyrG 215
         ||||||||||:::||||||:::|||||||||||||||:::      ||||
     861 TCAAAAGTGGGTAAATGGTATATCTTGGGGAATAGTGTAC......TATG 904

     215 lyGlyGlyAla.........GlySerThrLeuThrIleArgLeuArgIle 228
         |||||::::::         ||||||:::|||||||||||||||||||
```

```
     905 GAGGCTCTGGGAGAAAGAAAGGATCTGTTCTGACTATTCGCCTCAGAATA 954

     229 GluThrGlyThrGluProProValAlaMetGlyProAspLysValLeuAl 245
         ||||||      ||||||||||||||:::||||||:::|||   |||||
     955 GAAACTCAGATGGAACCTCCGGTTGCTATAGGACCAAATAAGGGTTTGGC 1004

     245 aGluGlnGlyProProAlaLeuGluProProHisAsnLeuProValProG 262
         ||||||||||||||||      |||
    1005 CGAACAAGGACCTCCAATCCAAGAACAG..................... 1032

     262 lnLeuThrSerLeuArgProAspIleThrGlnProProSerAsnSerThr 278
                     ||||||:::   :::            |||:::|||
    1033 ..............AGGCCATCTCCTAACCCCTCTGATTACAATACAACC 1068

     279 ThrGlyLeuIleProThrAsnThrProArgAsnSerProGlyValProVa 295
         :::|||   :::||||||                 :::|||::::::   ::
    1069 TCTGGATCAGTCCCCACT..............GAGCCTAACATCACTAT 1103

     295 lLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaI 312
         :|||||||||   :::||||||||||||||||||||||||||||||||||:
    1104 TAAAACAGGGGCGAAACTTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTC 1153

     312 leAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSer 328
         ::|||||||||   |||:::||||||||||||||||||||||||||:::
    1154 TTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCT 1203

     329 SerGlyProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnVa 345
         ||||||||||||||||||||||||||||:::      ||||||||||
    1204 TCGGGCCCACCTTACTATGAGGGAATGGCTAGAGGAGGGAAATTCAATGT 1253

     345 lThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuT 362
         ||||||||||||||||:::||||||||||||||||:::||||||||||
    1254 GACAAAGGAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTA 1303

     362 hrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProPro 378
         |||||||||||||||||||||||||||||||||||   :::||||||
    1304 CCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGGATGGTTCCCCCA 1353

     379 SerHisGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGl 395
         ||||||||||||||||||      |||   ::::::::::::::|||||
    1354 TCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGA 1403

     395 uAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrG 412
         |:::||||||||||||||||||||:::|||||||||||||||||||||||
    1404 GAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTG 1453

     412 lyLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspLeu 428
         ||||||||||||||||||||||   ||||||||||||:::||||||:::
    1454 GATTAACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAAAGACTTT 1503

     429 CysValMetValGlnIleValProArgValTyrTyrHisProGluGluVa 445
         ||||||||||||||||||||||||||||||||||||:::||||||:::::
    1504 TGCGTTATGGTCCAAATTGTCCCCCGGGTGTACTACTATCCCGAAAAAGC 1553

     445 lValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProV 462
         :||||||||||||||||||||||||||||||||||||||||||||||||:
    1554 AGTCCTTGATGAATATGACTATAGATATAATCGGCCAAAAAGAGAGCCCA 1603

     462 alSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyVal 478
         ::|||||||||||||||||||||||||||||||:::|||:::||||||
    1604 TATCCCTGACACTAGCTGTAATGCTCGGATTGGGAGTGGCTGCAGGCGTG 1653

     479 GlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGl 495
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1654 GGAACAGGAACGGCTGCCCTAATCACAGGACCGCAACAGCTGGAGAAAGG 1703

     495 yLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysG 512
         ||||::::::||||||      :::|||||||||||:::||||||::::
    1704 ACTTAGTAACCTACATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAA 1753

     512 luSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValVal 528
         ::|||||||||||||||||||||||||||||||||||||||||||||||||
    1754 AATCTGTCAGTAACCTGGAGGAATCCCTAACCTCCTTATCTGAAGTGGTT 1803
```

```
      529 LeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLe 545
          ||||||||||||||||||||||||||||||||||||||:::||||||||||
     1804 CTACAGAACAGAAGGGGGTTAGATCTGTTATTTCTAAAAGAAGGAGGGTT 1853

      545 uCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyA 562
          ||||:::||||||||||||||||||||||||||||||||||||||||||||
     1854 ATGTGTAGCCTTAAAAGAGGAATGCTGCTTCTATGTAGATCACTCAGGAG 1903

      562 laIleArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArg 578
          ||||||||||||||||:::||||||||:::::|||||||||||||||||
     1904 CCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGA 1953

      579 ArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSe 595
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1954 AGGGAAAGAGAGGCTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTC 2003

      595 rProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValL 612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     2004 TCCTTGGATGACCACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCC 2053

      612 euLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAla 628
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     2054 TGCTCCTGTTACTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCC 2103

      629 PheValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGl 645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     2104 TTTGTTAGAGAACGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACA 2153

      645 nTyrGlnGlyLeuLeuSerGlnGlyGluThrAspLeu 657
          |||||||||||||||||||||||||||||||||||||
     2154 GTACCAAGGCCTTCTGAGCCAAGGAGAAACTGACCTC 2190

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-2

seq_documentation_block:
; Sequence 2, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-2


alignment_scores:
           Quality: 2537.50              Length:    658
             Ratio:  4.452                Gaps:      9
 Percent Similarity:  86.626   Percent Identity: 72.796

alignment_block:
 US-09-171-553B-10 x US-08-766-528-2   ..

 Align seg 1/1  to: US-08-766-528-2  from: 1  to: 7333

       22 IleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrIleTh 38
          ||||||||||||||||||||||||||||||||||||||||||||:::|||||
     4806 ATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTCAATAAC 4855

       38 rProGlnAlaSerSerLysArgLeuIleAspSerSerAsnProHisArgP 55
          |||||||:::::::::|||||||||:::||||||   |||   |||:::|
     4856 TCCTCAAGTTAATGGTAAACGCCTTGTGGACAGCCCGAACTCCCATAAAC 4905

       55 roLeuSerProThrTrpLeuIleIleAspProAspThrGlyValThrVal 71
          ||||||||   |||||||||:::   |||      ||||||:::::::::
     4906 CCTTATCTCTCACCTGGTTACTTACTGACTCCGGTACAGGTATTAATATT 4955

       72 AsnSerThrArgGlyValAlaProArgGlyThrTrpTrpProGluLeuHi 88
          |||||||||:::|||   ||||||   ||||||||||||||||||||::
     4956 AACAGCACTCAAGGGGAGGCTCCCTTGGGGACCTGGTGGCCTGAATTATA 5005

       88 sPheCysLeuArgLeuIleAsnProAlaVal......LysSerThrProP 103
          :   |||||||||   :::   |||::::::      ::::::|||||||
     5006 TGTCTGCCTTCGATCAGTAATCCCTGGTCTCAATGACCAGGCCACACCCC 5055

      103 roAsnLeuValArgSerTyrGlyPheTyrCysCysProGly...ThrGlu 118
          ||:::::::::|||:::||||||||||||   |||||||||      :::
     5056 CCGATGTACTCCGTGCTTACGGGTTTTACGTTTGCCCAGGACCCCCAAAT 5105

      119 LysGluLysTyrCysGlyGlySerGlyGluSerPheCysArgArgTrpSe 135
          :::|||:::|||||||||:::      :::   ||||||::::::|||||
     5106 AATGAAGAATATTGTGGAAATCCTCAGGATTTCTTTTGCAAGCAATGGAG 5155

      135 rCysValThrSerAsnAspGlyAspTrpLysTrpProIleSerLeuGlnA 152
          ||||:::||||||||||||||:::||||||||||||:::|||   ||||
     5156 CTGCATAACTTCTAATGATGGGAATTGGAAATGGCCAGTCTCTCAGCAAG 5205

      152 spArgValLysPheSerPheValAsnSer..................... 161
          ||||||||::::::|||||||||||||:::
     5206 ACAGAGTAAGTTACTCTTTTGTTAACAATCCTACCAGTTATAATCAATTT 5255

      162 ......GlyProGlyLysTyrLys......MetMetLysLeuTyrLysAs 174
                ||||   |||::::::|||      ::::::::::::   |||||
     5256 AATTATGGCCATGGGAGATGGAAAGATTGGCAACAGCGGGTACAAAAAGA 5305

      174 pLys..............SerCysSerProSerAspLeuAspTyrLeuL 186
          |                |||||         ||||||||||||||||
     5306 TGTACGAAATAAGCAAATAAGCTGTCATTCGTTAGACCTAGATTACTTAA 5355

      186 ysIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrpIle 202
          ||||||||||||||||||||||||||||||||||||||||||||||||:::
     5356 AAATAAGTTTCACTGAAAAAGGAAAACAAGAAAATATTCAAAAGTGGGTA 5405

      203 AsnGlyMetSerTrpGlyIleValPheTyrLysTyrGlyGlyGlyAla.. 218
          ||||||:::|||||||||||||||:::      |||||||||::::::
     5406 AATGGTATATCTTGGGGAATAGTGTAC......TATGGAGGCTCTGGGAG 5449

      219 .......GlySerThrLeuThrIleArgLeuArgIleGluThrGlyThrG 233
                ||||||:::|||||||||||||||||||||||||||      |
     5450 AAAGAAAGGATCTGTTCTGACTATTCGCCTCAGAATAGAAACTCAGATGG 5499

      233 luProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyPro 249
          ||||||||||||||:::|||||||:::|||   ||||||||||||||||||
```

```
5500 AACCTCCGGTTGCTATAGGACCAAATAAGGGTTTGGCCGAACAAGGACCT 5549

 250 ProAlaLeuGluProProHisAsnLeuProValProGlnLeuThrSerLe 266
     |||      |||
5550 CCAATCCAAGAACAG..................................... 5564

 266 uArgProAspIleThrGlnProProSerAsnSerThrThrGlyLeuIleP 283
      |||||||:::   :::             |||:::|||:::|||   :::|
5565 .AGGCCATCTCCTAACCCCTCTGATTACAATACAACCTCTGGATCAGTCC 5613

 283 roThrAsnThrProArgAsnSerProGlyValProValLysThrGlyGln 299
     |||||               :::|||:::::   :::|||||||||
5614 CCACT..............GAGCCTAACATCACTATTAAAACAGGGGCG 5648

 300 ArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSerThrAs 316
     :::||||||||||||||||||||||||||||||||||:::|||||||||
5649 AAACTTTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACGAC 5698

 316 pProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProProT 333
      |||:::|||||||||||||||||||||||||||||:::|||||||||||
5699 TCCAGAGGCTACCTCTTCTTGTTGGCTTTGCTTAGCTTCGGGCCCACCTT 5748

 333 yrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGluHis 349
     ||||||||||||||||:::      |||||||||||||||||||||||||
5749 ACTATGAGGGAATGGCTAGAGGAGGGAAATTCAATGTGACAAAGGAACAT 5798

 350 ArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThrGluVa 366
     |||:::|||||||||||||||||||:::|||||||||||||||||||||
5799 AGAGACCAATGTACATGGGGATCCCAAAATAAGCTTACCCTTACTGAGGT 5848

 366 lSerGlyLysGlyThrCysIleGlyLysAlaProProSerHisGlnHisL 383
     |||||||||||||||||||||||||   :::||||||||||||||||||
5849 TTCTGGAAAAGGCACCTGCATAGGGATGGTTCCCCCATCCCACCAACACC 5898

 383 euCysTyrSerThrValValTyrGluGlnAlaSerGluAsnGlnTyrLeu 399
     |||||      |||   ::::::::::::::|||||:::|||||||||
5899 TTTGTAACCACACTGAAGCCTTTAATCGAACCTCTGAGAGTCAATATCTG 5948

 400 ValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLeuThrProCy 416
     ||||||||||||:::|||||||||||||||||||||||||||||||||||
5949 GTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATTAACCCCTTG 5998

 416 sValSerThrSerValPheAsnGlnSerLysAspLeuCysValMetValG 433
     ||||||||||   ||||||||||||:::|||||||:::|||||||||||||
5999 TGTTTCCACCTTGGTTTTCAACCAAACTAAAGACTTTTGCGTTATGGTCC 6048

 433 lnIleValProArgValTyrTyrHisProGluGluValValLeuAspGlu 449
     |||||||||||||||||||||||::::||||||::::::|||||||||||
6049 AAATTGTCCCCCGGGTGTACTACTATCCCGAAAAAGCAGTCCTTGATGAA 6098

 450 TyrAspTyrArgTyrAsnArgProLysArgGluProValSerLeuThrLe 466
     |||||||||||||||||||||||||||||||||||:::|||||||||||
6099 TATGACTATAGATATAATCGGCCAAAAAGAGAGCCCATATCCCTGACACT 6148

 466 uAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGlyThrA 483
     |||||||||||||||||||||||:::|||:::|||||||||||||||||
6149 AGCTGTAATGCTCGGATTGGGAGTGGCTGCAGGCGTGGGAACAGGAACGG 6198

 483 laAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeu 499
     |||||||||||||||||||||||||||||||||||||||:::::::|||
6199 CTGCCCTAATCACAGGACCGCAACAGCTGGAGAAAGGACTTAGTAACCTA 6248

 500 HisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAs 516
     |||      :::|||||||||||||:::||||||:::::|||||||||||
6249 CATCGAATTGTAACGGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAA 6298

 516 nLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgA 533
     |||||||||||||||||||||||||||||||||||||||||||||||||
6299 CCTGGAGGAATCCCTAACCTCCTTATCTGAAGTGGTTCTACAGAACAGAA 6348

 533 rgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeu 549
     |||||||||||||||||||||||:::|||||||||||||||:::||||||
6349 GGGGGTTAGATCTGTTATTTCTAAAAGAAGGAGGGTTATGTGTAGCCTTA 6398

 550 LysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSe 566
     ||||||||||||||||||||||||||||||||||||||||||||||||||
6399 AAAGAGGAATGCTGCTTCTATGTAGATCACTCAGGAGCCATCAGAGACTC 6448

 566 rMetAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArgGluA 583
     ||||:::|||||||||:::::|||||||||||||||||||||||||||||
6449 CATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCGAAGGGAAAGAGAGG 6498

 583 laAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMetThr 599
     ||||||||||||||||||||||||||||||||||||||||||||||||||
6499 CTGACCAGGGGTGGTTTGAAGGATGGTTCAACAGGTCTCCTTGGATGACC 6548

 600 ThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLe 616
     ||||||||||||||||||||||||||||||||||||||||||||||||||
6549 ACCCTGCTTTCTGCTCTGACGGGGCCCCTAGTAGTCCTGCTCCTGTTACT 6598

 616 uThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluA 633
     ||||||||||||||||||||||||||||||||||||||||||||||||||
6599 TACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTGCCTTTGTTAGAGAAC 6648

 633 rgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 649
     ||||||||||||||||||||||||||||||||||||||||||||||||||
6649 GAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAGGCCTT 6698

 650 LeuSerGlnGlyGluThrAspLeu 657
     |||||||||||||||||||||||
6699 CTGAGCCAAGGAGAAACTGACCTC 6722
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-1

```
seq_documentation_block:
; Sequence 1, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1
```

```
alignment_scores:
           Quality: 2402.00            Length:    655
             Ratio:  4.312              Gaps:      9
 Percent Similarity:  85.038   Percent Identity: 68.702

alignment_block:
 US-09-171-553B-10 x US-08-766-528-1   ..

 Align seg 1/1  to: US-08-766-528-1  from: 1  to: 8060

       1 MetHisProThrLeuSerTrpArgHisLeuProThrArgGlyGlyGluPr 17
      86 ATGCATCCCACGTTAAGCCGGCGCCACCTCCCGATTCGGGGTGGAAAGCC 135

      17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
     136 GAAAAGACTGAAAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTA 185

      34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
     186 CTCTGTCAATAACCTCTCAGACTAATGGTATGCGCATAGGAGACAGCCTG 235

      51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAspTh 67
     236 AACTCCCATAAACCCTTATCTCTCACCTGGTTAATTACTGACTCCGGCAC 285

      67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
     286 AGGTATTAATATCAACAACACTCAAGGGGAGGCTCCTTTAGGAACCTGGT 335

      84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
     336 GGCCTGATCTATACGTTTGCCTCAGATCAGTTATTCCTAGTCTGACCTCA 385

     101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyTh 117
     386 ...CCCCCAGATATCCTCCATGCTCACGGATTTTATGTTTGCCCAGGACC 432

     117 rGluLysGlu...LysTyrCysGlyGlySerGlyGluSerPheCysArgA 133
     433 ACCAAATAATGGAAAACATTGCGGAAATCCCAGAGATTTCTTTTGTAAAC 482

     133 rgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149
     483 AATGGAACTGTGTAACCTCTAATGATGGATATTGGAAATGGCCAACCTCT 532

     150 LeuGlnAspArgValLysPheSerPheValAsn......SerGlyProGl 164
     533 CAGCAGGATAGGGTAAGTTTTTCTTATGTCAACACCTATACCAGCTCTGG 582

     164 yLysTyrLysMetMetLysLeuTyrLysAspLysSer......CysSerP 179
     583 ACAATTTAATTACCTGACCTGGATTAGAACTGGAAGCCCCAAGTGCTCTC 632

     179 roSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGln 195
     633 CTTCAGACCTAGATTACCTAAAAATAAGTTTCACTGAGAAAGGAAAACAA 682

     196 GluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTy 212
     683 GAAAATATCCTAAAATGGGTAAATGGTATGTCTTGGGGAATGGTATAT.. 730

     212 rLysTyrGlyGlyGlyAla.........GlySerThrLeuThrIleArgL 226
     731 ....TATGGAGGCTCGGGTAAACAACCAGGCTCCATTCTAACTATTCGCC 776

     226 euArgIleGluThrGlyThrGluProProValAlaMetGlyProAspLys 242
     777 TCAAAATAAACCAG...CTGGAGCCTCCAATGGCTATAGGACCAAATACG 823

     243 ValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPr 259

     824 GTCTTGACGGGTCAAAGACCCCCAACCCAAGGACCA..............  859

     259 oValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerA 276
     860 ..................................GGACCATCCTCTA 872

     276 snSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 292
     873 ACATAACTTCTGGATCAGACCCCACTGAGTCTAGCAGCACGACT...... 916

     293 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPh 309
     917 .........AAAATGGGGGCAAAACTTTTTAGCCTCATCCAGGGAGCTTT 957

     309 eGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLeuC 326
     958 TCAAGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTAT 1007

     326 ysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluArgLys 342
    1008 GCTTAGCTTTGGGCCCACCTTACTATGAAGGAATGGCTAGAAGAGGGAAA 1057

     343 PheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAs 359
    1058 TTCAATGTGACAAAAGAACATAGAGACCAATGCACATGGGGATCCCAAAA 1107

     359 nLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysA 376
    1108 TAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGAAAGG 1157

     376 laProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 392
    1158 TTCCCCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCAA 1207

     393 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCy 409
    1208 ACCTCTGAAAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATG 1257

     409 sAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSerL 426
    1258 TAATACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTTAACCAAACTA 1307

     426 ysAspLeuCysValMetValGlnIleValProArgValTyrTyrHisPro 442
    1308 AAGATTTTTGCATTATGGTCCAAATTGTTCCCCGAGTGTATTACTATCCC 1357

     443 GluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgProLysAr 459
    1358 GAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAAGAG 1407

     459 gGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaV 476
    1408 AGAACCCATATCTCTGACACTTGCTGTGATGCTCGGACTTGGAGTGGCAG 1457

     476 alGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 492
    1458 CAGGTGTAGGAACAGGAACAGCTGCCCTGGTCACGGGACCACAGCAGCTA 1507

     493 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAl 509
    1508 GAAACAGGACTTAGTAACCTACATCGAATTGTAACAGAAGATCTCCAAGC 1557

     509 aLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSerG 526
    1558 CCTAGAAAAATCTGTCAGTAACCTGGAGGAATCCCTAACCTCCTTATCTG 1607

     526 luValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGlu 542
    1608 AAGTAGTCCTACAGAATAGAAGAGGGTTAGATTTATTATTTCTAAAAGAA 1657

     543 GlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHi 559
    1658 GGAGGATTATGTGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTGGATCA 1707
```

```
      559 sSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGluA 576
          ||||||||||||||||||||||||||||||||||||||:::::|||||||:
     1708 TTCAGGGGCCATCAGAGACTCCATGAACAAACTTAGAGAAAGGTTGGAGA 1757

      576 rgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPhe 592
          ::|||||||||||:::|||:::   |||||||||||||||||||||||||
     1758 AGCGTCGAAGGGAAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGGTTC 1807

      593 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLe 609
          ||||||||||||||:::::::||||||||||||||||||||||||||||||
     1808 AACAGGTCTCCTTGGTTGGCTACCCTACTTTCTGCTTTAACAGGACCCTT 1857

      609 uValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgP 626
          |:::|||||||||||||||||||||||||||||||:::|||||||::::
     1858 AATAGTCCTCCTCCTGTTACTCACAGTTGGGCCATGTATTATTAACAAGT 1907

      626 heValAlaPheValArgGluArgValSerAlaValGlnIleMetValLeu 642
          :::::||||||:::|||||||||:::|||||||||||||||||||||||||
     1908 TAATTGCCTTCATTAGAGAACGAATAAGTGCAGTCCAGATCATGGTACTT 1957

      643 ArgGlnGlnTyrGln 647
          |||||||||||||||
     1958 AGACAACAGTACCAA 1972

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-3

seq_documentation_block:
; Sequence 3, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8132 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-3


alignment_scores:
            Quality: 2395.00          Length:    655
              Ratio:  4.292            Gaps:     10

 Percent Similarity:  85.191   Percent Identity: 68.855

alignment_block:
 US-09-171-553B-10 x US-08-766-528-3   ..

 Align seg 1/1  to: US-08-766-528-3  from: 1  to: 8132

        1 MetHisProThrLeuSerTrpArgHisLeuProThrArgGlyGlyGluPr 17
          |||||||||||||||:::   ||||||||||||   ||||||||||:::||
     5620 ATGCATCCCACGTTAAACCGGCGCCACCTCCCGATTCGGGGTGGAAAGCC 5669

       17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
          ||||||||||:::|||||||||||||||||||||||||||||||||||||
     5670 GAAAAGACTGAAAATCCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTA 5719

       34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
          ||||||:::||||||   |||:::::::::   |||:::   ||||||
     5720 CTCTGTCAATAACCTCTCAGACTAATGGTATGCGCATAGGAGACAGCCTG 5769

       51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAspTh 67
          |||   |||:::|||||||||   ||||||||||||   |||       ||
     5770 AACTCCCATAAACCCTTATCTCTCACCTGGTTAATTACTGACTCCGGCAC 5819

       67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
          ||||:::::::::|||:::|||:::|||   ||||||   |||||||||
     5820 AGGTATTAATATCAACAACACTCAAGGGGAGGCTCCTTTAGGAACCTGGT 5869

       84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
          |||||:::|||:::   |||||||||   :::   |||:::::::   |||
     5870 GGCCTGATCTATACGTTTGCCTCAGATCAGTTATTCCTAGTCTGACCTCA 5919

      101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyTh 117
          ||||||:::::::::::::::::::::|||||||||   |||||||||
     5920 ...CCCCCAGATATCCTCCATGCTCACGGATTTTATGTTTGCCCAGGACC 5966

      117 rGluLysGlu...LysTyrCysGlyGlySerGlyGluSerPheCysArgA 133
            ::::::   |||:::||||||:::       :::   ||||||::::
     5967 ACCAAATAATGGAAAACATTGCGGAAATCCCAGAGATTTCTTTTGTAAAC 6016

      133 rgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149
          ::|||:::||||||||||||||||||||   ||||||||||||   |||
     6017 AATGGAACTGTGTAACCTCTAATGATGGATATTGGAAATGGCCAACCTCT 6066

      150 LeuGlnAspArgValLysPheSerPheValAsn......SerGlyProGl 164
            ||||||||||||:::||||||:::||||||      ::::::   ||
     6067 CAGCAGGATAGGGTAAGTTTTTCTTATGTCAACACCTATACCAGCTCTGG 6116

      164 yLysTyrLysMetMetLysLeuTyrLysAspLysSer......CysSerP 179
          |:::::::::   :::         :::      |||      |||||||
     6117 ACAATTTAATTACCTGACCTGGATTAGAACTGGAAGCCCCAAGTGCTCTC 6166

      179 roSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGln 195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     6167 CTTCAGACCTAGATTACCTAAAAATAAGTTTCACTGAGAAAGGAAAACAA 6216

      196 GluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTy 212
          |||||||||   |||||:::|||||||||||||||||||:::|||:::
     6217 GAAAATATCCTAAAATGGGTAAATGGTATGTCTTGGGGAATGGTATAT.. 6264

      212 rLysTyrGlyGlyGlyAla.........GlySerThrLeuThrIleArgL 226
            |||||||||::::::         ||||||   ||||||||||||||
     6265 ....TATGGAGGCTCGGGTAAACAACCAGGCTCCATTCTAACTATTCGCC 6310

      226 euArgIleGluThrGlyThrGluProProValAlaMetGlyProAspLys 242
          ||:::|||:::         ||||||||||:::|||:::||||||:::
     6311 TCAAAATAAACCAG...CTGGAGCCTCCAATGGCTATAGGACCAAATACG 6357

      243 ValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPr 259
          |||||||:::   |||   ||||||:::      |||
     6358 GTCTTGACGGGTCAAAGACCCCCAACCCAAGGACCA.............. 6393

      259 oValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerA 276
                                                  |||   ||||
```

```
6394 .....GGACCATCCTCTA 6406

 276 snSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 292
     ||   |||:::|||      ||||||:::::      ||||||
6407 ACATAACTTCTGGATCAGACCCCACTGAGTCT......AACAGC...... 6444

 293 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPh 309
         :::|||   |||   :::||||||||||||||||||||||||||||
6445 ...ACGACTAAAATGGGGGCAAAACTTTTTAGCCTCATCCAGGGAGCTTT 6491

 309 eGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLeuC 326
     ||||||||:::|||||||||   |||:::||||||||||||||||||||||
6492 TCAAGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTAT 6541

 326 ysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluArgLys 342
     |||||:::||||||||||||||||||||||||||||||::::::   |||
6542 GCTTAGCTTCGGGCCCACCTTACTATGAAGGAATGGCTAGAAGAGGGAAA 6591

 343 PheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAs 359
     ||||||||||||||||||||||||:::||||||||||||||||||:::||
6592 TTCAATGTGACAAAAGAACATAGAGACCAATGCACATGGGGATCCCAAAA 6641

 359 nLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysA 376
     ||||||||||||||||||||||||||||||||||||||||||||||||:
6642 TAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGAAAGG 6691

 376 laProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 392
     ::|||||||||||||||||||||||      |||   ::::::::::|||
6692 TTCCCCCATCCCACCAACACCTTTGTAACCACACTGAAGCCTTTAATCAA 6741

 393 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCy 409
     :::||||||:::|||||||||||||||||||:::||||||||||||||||
6742 ACCTCTGAGAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATG 6791

 409 sAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSerL 426
     |||||||||||||||||||||||||||||   ||||||||||||:::|
6792 TAATACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTTAACCAAACTA 6841

 426 ysAspLeuCysValMetValGlnIleValProArgValTyrTyrHisPro 442
     |||||:::|||:::|||||||||||||||||||||||||||||:::|||
6842 AAGATTTTTGCATTATGGTCCAAATTGTTCCCCGAGTGTATTACTATCCC 6891

 443 GluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgProLysAr 459
     |||:::::::::|||||||||||||||||||||   :::|||   |||||
6892 GAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAAGAG 6941

 459 gGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaV 476
     |||||||:::|||||||||||||||||||||||||||||||||:::|||:
6942 AGAACCCATATCTCTGACACTTGCTGTGATGCTCGGACTTGGAGTGGCAG 6991

 476 alGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 492
     ::|||||||||||||||||||||||||||||:::||||||||||||||||
6992 CAGGTGTAGGAACAGGAACAGCTGCCCTGGTCACGGGACCACAGCAGCTA 7041

 493 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAl 509
     |||   ||||||:::::||||||      :::|||||||||||||:::||
7042 GAAACAGGACTTAGTAACCTACATCGAATTGTAACAGAAGATCTCCAAGC 7091

 509 aLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSerG 526
     ||||::::::|||||||||||||||||||||||||||||||||||||||
7092 CCTAGAAAAATCTGTCAGTAACCTGGAGGAATCCCTAACCTCCTTATCTG 7141

 526 luValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGlu 542
     ||||||||||||||||||||||||||||||||||||||||||||:::|||
7142 AAGTAGTCCTACAGAATAGAAGAGGGTTAGATTTATTATTTCTAAAAGAA 7191

 543 GlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHi 559
     ||||||||||||:::|||||||||||||||||||||||||||||||||||
7192 GGAGGATTATGTGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTGGATCA 7241

 559 sSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGluA 576
     ||||||||||||||||||||||||||||||||||:::::::|||||||:
7242 TTCAGGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAAGGTTGGAGA 7291

 576 rgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPhe 592
     ::|||||||||||||:::|||:::   ||||||||||||||||||||||||
7292 AGCGTCGAAGGGAAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGGTTC 7341

 593 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLe 609
     |||||||||   |||:::::::||||||||||||||||||||||||||||
7342 AACAGGTCTCTTTGGTTGGCTACCCTACTTTCTGCTTTAACAGGACCCTT 7391

 609 uValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgP 626
     |:::|||||||||||||||||||||||||||||||||:::||||||::::
7392 AATAGTCCTCCTCCTGTTACTCACAGTTGGGCCATGTATTATTAACAAGT 7441

 626 heValAlaPheValArgGluArgValSerAlaValGlnIleMetValLeu 642
     :::::||||||:::||||||||||:::||||||||||||||||||||||||
7442 TAATTGCCTTCATTAGAGAACGAATAAGTGCAGTCCAGATCATGGTACTT 7491

 643 ArgGlnGlnTyrGln 647
     |||||||||||||||
7492 AGACAACAGTACCAA 7506
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-25

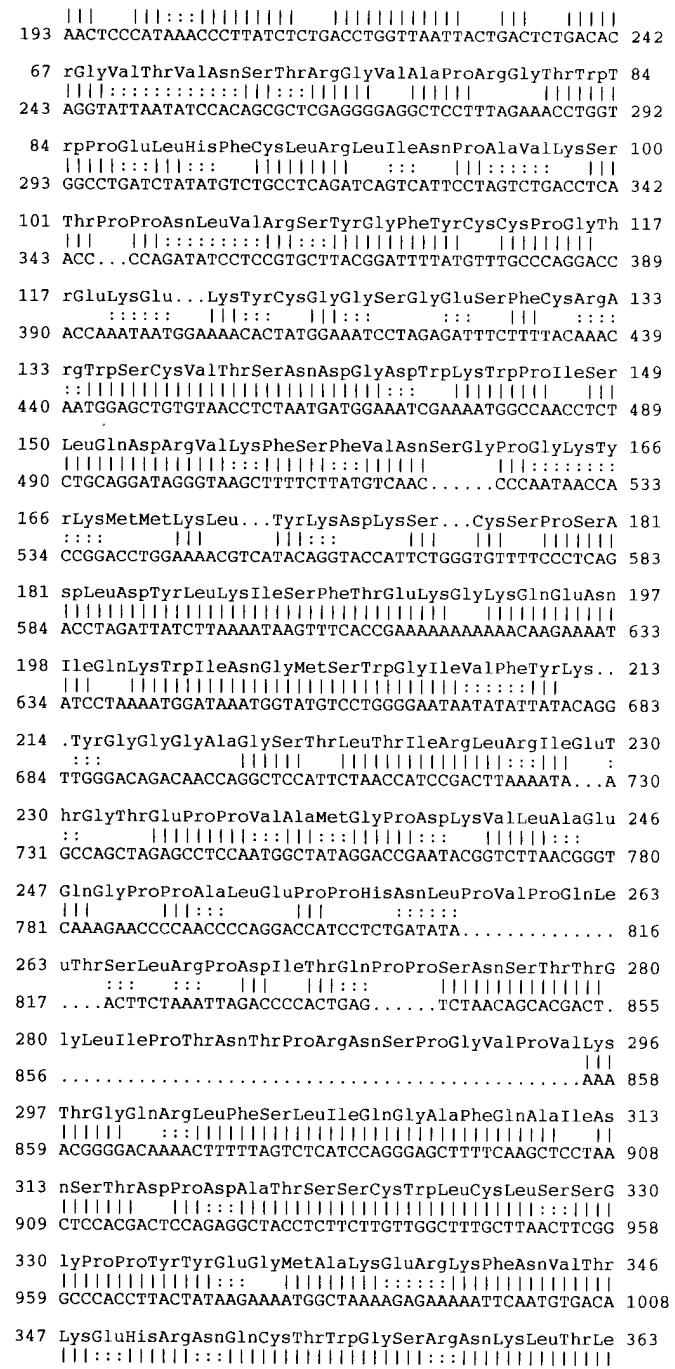
seq_documentation_block:
; Sequence 25, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Contiguous
; OTHER INFORMATION: compilation of SEQ ID NOs: 18, 22, and 24.
US-09-376-781-25

alignment_scores:
Quality: 2169.00 Length: 652
Ratio: 4.032 Gaps: 11
Percent Similarity: 82.515 Percent Identity: 64.571

alignment_block:
US-09-171-553B-10 x US-09-376-781-25 ..

Align seg 1/1 to: US-09-376-781-25 from: 1 to: 2000

```
   1 MetHisProThrLeuSerTrpArgHisLeuProThrArgGlyGlyGluPr 17
     :::|||||||||||||||   |||||||||      :::|||   :::||
  43 ATACATCCCACGTTAAGCCGGCGCCACCTCCTGATTCAGGGTGAAAAGCC 92

  17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
     ||||||||||:::|||||||||||||||||||||||:::||||||||||
  93 AAAAAGACTAAAAATCCCCTTAAGCTTCGCCTCCATCACATGGTTCCTTA 142

  34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
     |||||:::||||||   |||:::::::::   ::::::   ||||||
 143 CTCTGTCAATAACCTCTCAGACTAATGGTATGCACATAGGAGACAGCCTG 192

  51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAspTh 67
```


193 AACTCCCATAAACCCTTATCTCTGACCTGGTTAATTACTGACTCTGACAC 242
67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
243 AGGTATTAATATCCACAGCGCTCGAGGGGAGGCTCCTTTAGAAACCTGGT 292
84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
293 GGCCTGATCTATATGTCTGCCTCAGATCAGTCATTCCTAGTCTGACCTCA 342
101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyTh 117
343 ACC...CCAGATATCCTCCGTGCTTACGGATTTTATGTTTGCCCAGGACC 389
117 rGluLysGlu...LysTyrCysGlyGlySerGlyGluSerPheCysArgA 133
390 ACCAAATAATGGAAAACACTATGGAAATCCTAGAGATTTCTTTTACAAAC 439
133 rgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149
440 AATGGAGCTGTGTAACCTCTAATGATGGAAATCGAAAATGGCCAACCTCT 489
150 LeuGlnAspArgValLysPheSerPheValAsnSerGlyProGlyLysTy 166
490 CTGCAGGATAGGGTAAGCTTTTCTTATGTCAAC......CCCAATAACCA 533
166 rLysMetMetLysLeu...TyrLysAspLysSer...CysSerProSerA 181
534 CCGGACCTGGAAAACGTCATACAGGTACCATTCTGGGTGTTTTCCCTCAG 583
181 spLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsn 197
584 ACCTAGATTATCTTAAAATAAGTTTCACCGAAAAAAAAAAACAAGAAAAT 633
198 IleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTyrLys.. 213
634 ATCCTAAAATGGATAAATGGTATGTCCTGGGGAATAATATATTATACAGG 683
214 .TyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGluT 230
684 TTGGGACAGACAACCAGGCTCCATTCTAACCATCCGACTTAAAATA...A 730
230 hrGlyThrGluProProValAlaMetGlyProAspLysValLeuAlaGlu 246
731 GCCAGCTAGAGCCTCCAATGGCTATAGGACCGAATACGGTCTTAACGGGT 780
247 GlnGlyProProAlaLeuGluProProHisAsnLeuProValProGlnLe 263
781 CAAAGAACCCCAACCCCAGGACCATCCTCTGATATA..... 816
263 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrG 280
817 ....ACTTCTAAATTAGACCCCACTGAG......TCTAACAGCACGACT. 855
280 lyLeuIleProThrAsnThrProArgAsnSerProGlyValProValLys 296
856 .....AAA 858
297 ThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAs 313
859 ACGGGGACAAAACTTTTTAGTCTCATCCAGGGAGCTTTTCAAGCTCCTAA 908
313 nSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerG 330
909 CTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTTGCTTAACTTCGG 958
330 lyProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThr 346
959 GCCCACCTTACTATAAGAAAATGGCTAAAAGAGAAAAATTCAATGTGACA 1008
347 LysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLe 363
1009 AAAAAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTACCCT 1058
363 uThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProProSerH 380
1059 TACTGAGGTTTCTGGAAAGGACACCTGCATAAAAAAGGTTCCCCCATCCC 1108
380 isGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsn 396
1109 ACCAACACCTTTACAACCACACTGAAGCCTTTAATCAAACCTCTGAGAGT 1158
397 GlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLe 413
1159 CAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATT 1208
413 uThrProCysValSerThrSerValPheAsnGlnSerLysAspLeuCysV 430
1209 AACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAAAGACTTTTACA 1258
430 alMetValGlnIleValProArgValTyrTyrHisProGluGluValVal 446
1259 TTATGGTCCAAATTGTTCCCCGAGTATATTACTATCCCAAGAAAACAATT 1308
447 LeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSe 463
1309 CTCGATGAATATGATTACAGGAACCATCGACAAAAGAAAAAACCCATATC 1358
463 rLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyT 480
1359 CCTGACACTCGCAGTAATGCTCGGACTCGGAGTGATAACAGGTGTGAGAA 1408
480 hrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeu 496
1409 CAGGAACTGCAGCTTTAGTTACAGGACCTCAGCAGCTAGAAACAGGACTT 1458
497 GlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSe 513
1459 AGTAACCTACATCAAATTGTAACAGGAAATCTCCAAGCCCTAAAAAAATC 1508
513 rValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuG 530
1509 TGTCAGTAACCTGGAAAAATCCCTAACCTCCTTATCTGAAGTAGTTCTAC 1558
530 lnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCys 546
1559 AGAATAAAAAAGGGTTAGATTTATTATTTCTAAAAAAAAGAAGATTATGT 1608
547 AlaAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyAlaIl 563
1609 GTAGCCTTAAAGGAGAAATGCTGTTTTTATGTAGATCATTCAGGGGCCAT 1658
563 eArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArgArgG 580
1659 CAGAGACTCCATGAACAAGCTTAAAAAAAGGTTGGAGAAACGTCGAAGGG 1708
580 luArgGluAla.AspGlnGlyTrpPheGluGlyTrpPheAsnArgSerPr 596
1709 AAAAGGAAACTTACTCAAAGATGGTTTAAAAGATGGTTCAACAGGTCTCC 1758
596 oTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuL 613
1759 TTGGTTGGCTACCCTACTTTCTACTTTAACAGGACCCTTAATAGTCCTCC 1808
613 euLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPhe 629
1809 TCCTGTTACTCACAGTTGGGCCATGTATTATTAACAAGTTAATTGCCTTC 1858
630 ValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGlnTy 646
1859 ATTAGAAAACGAATAAGTGCAGTCCAGATCATGGTACTTAGACAACAGTA 1908
646 rGln 647
1909 CCAA 1912

```
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-30

seq_documentation_block:
; Sequence 30, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 2000
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: consensus sequence of PERV-D.
US-09-376-781-30


alignment_scores:
            Quality: 2169.00            Length:    652
              Ratio:  4.032              Gaps:     11
 Percent Similarity:  82.515   Percent Identity: 64.571

alignment_block:
 US-09-171-553B-10 x US-09-376-781-30   ..

 Align seq 1/1  to: US-09-376-781-30  from: 1  to: 2000

       1 MetHisProThrLeuSerTrpArgHisLeuProThrArgGlyGlyGluPr 17
         :::|||||||||||||||   |||||||||      :::|||   :::||
      43 ATACATCCCACGTTAAGCCGGCGCCACCTCCTGATTCAGGGTGAAAAGCC 92

      17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
         ||||||||||:::|||||||||||||||||||||||:::||||||||||
      93 AAAAAGACTAAAAATCCCCTTAAGCTTCGCCTCCATCACATGGTTCCTTA 142

      34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
         |||||:::||||||   |||:::::::::   ::::::   ||||||
     143 CTCTGTCAATAACCTCTCAGACTAATGGTATGCACATAGGAGACAGCCTG 192

      51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAspTh 67
         |||   |||:::|||||||||   ||||||||||||   |||   |||||
     193 AACTCCCATAAACCCTTATCTCTGACCTGGTTAATTACTGACTCTGACAC 242

      67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrpT 84
         ||||::::::::::::|||:::||||||   ||||||      |||||||
     243 AGGTATTAATATCCACAGCGCTCGAGGGGAGGCTCCTTTAGAAACCTGGT 292

      84 rpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
         |||||:::|||:::   |||||||||   :::   |||::::::   |||
     293 GGCCTGATCTATATGTCTGCCTCAGATCAGTCATTCCTAGTCTGACCTCA 342

     101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyTh 117
         |||   |||::::::::::|||:::||||||||||   |||||||||
     343 ACC...CCAGATATCCTCCGTGCTTACGGATTTTATGTTTGCCCAGGACC 389

     117 rGluLysGlu...LysTyrCysGlyGlySerGlyGluSerPheCysArgA 133
             ::::::   |||:::   |||:::      :::   |||   ::::
     390 ACCAAATAATGGAAAACACTATGGAAATCCTAGAGATTTCTTTTACAAAC 439

     133 rgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149

         ::|||||||||||||||||||||||||:::   |||||||||   |||
     440 AATGGAGCTGTGTAACCTCTAATGATGGAAATCGAAAATGGCCAACCTCT 489

     150 LeuGlnAspArgValLysPheSerPheValAsnSerGlyProGlyLysTy 166
         ||||||||||||||:::||||||:::||||||      |||:::::::::
     490 CTGCAGGATAGGGTAAGCTTTTCTTATGTCAAC......CCCAATAACCA 533

     166 rLysMetMetLysLeu...TyrLysAspLysSer...CysSerProSerA 181
         ::::      |||      |||:::      |||   |||   |||||||
     534 CCGGACCTGGAAAACGTCATACAGGTACCATTCTGGGTGTTTTCCCTCAG 583

     181 spLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsn 197
         |||||||||||||||||||||||||||||||||   ||||||||||||
     584 ACCTAGATTATCTTAAAATAAGTTTCACCGAAAAAAAAAAACAAGAAAAT 633

     198 IleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTyrLys.. 213
         |||   ||||||||||||||||||||||||||||||:::::::|||
     634 ATCCTAAAATGGATAAATGGTATGTCCTGGGGAATAATATATTATACAGG 683

     214 .TyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGluT 230
          :::            ||||||   ||||||||||||||:::|||   :
     684 TTGGGACAGACAACCAGGCTCCATTCTAACCATCCGACTTAAAATA...A 730

     230 hrGlyThrGluProProValAlaMetGlyProAspLysValLeuAlaGlu 246
         ::      |||||||||:::|||:::||||||:::   ||||||:::
     731 GCCAGCTAGAGCCTCCAATGGCTATAGGACCGAATACGGTCTTAACGGGT 780

     247 GlnGlyProProAlaLeuGluProProHisAsnLeuProValProGlnLe 263
         |||     |||:::      |||      ::::::
     781 CAAAGAACCCCAACCCCAGGACCATCCTCTGATATA.............. 816

     263 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrG 280
             :::   :::   |||   |||:::      ||||||||||||||||
     817 ....ACTTCTAAATTAGACCCCACTGAG......TCTAACAGCACGACT. 855

     280 lyLeuIleProThrAsnThrProArgAsnSerProGlyValProValLys 296
                                                         |||
     856 ..............................................AAA 858

     297 ThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAs 313
         ||||||   :::||||||||||||||||||||||||||||||||||   ||
     859 ACGGGGACAAAACTTTTTAGTCTCATCCAGGGAGCTTTTCAAGCTCCTAA 908

     313 nSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerG 330
         |||||||   |||:::|||||||||||||||||||||||||||:::||||
     909 CTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTTGCTTAACTTCGG 958

     330 lyProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThr 346
         ||||||||||||||::   |||||||||:::::|||||||||||||||||
     959 GCCCACCTTACTATAAGAAAATGGCTAAAAGAGAAAAATTCAATGTGACA 1008

     347 LysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLe 363
         |||:::||||||:::||||||||||||||||||||:::||||||||||||
    1009 AAAAAACATAGAGACCAATGTACATGGGGATCCCAAAATAAGCTTACCCT 1058

     363 uThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProProSerH 380
         ||||||||||||||||||   |||||||||   |||:::||||||||||||
    1059 TACTGAGGTTTCTGGAAAGGACACCTGCATAAAAAAGGTTCCCCCATCCC 1108

     380 isGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsn 396
         ||||||||||          |||   ::::::::::|||:::||||||:::
    1109 ACCAACACCTTTACAACCACACTGAAGCCTTTAATCAAACCTCTGAGAGT 1158

     397 GlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLe 413
         ||||||||||||||||||||:::||||||||||||||||||||||||||||
    1159 CAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACTGGATT 1208

     413 uThrProCysValSerThrSerValPheAsnGlnSerLysAspLeuCysV 430
         ||||||||||||||||||   ||||||||||||:::||||||:::   :
    1209 AACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAAAGACTTTTACA 1258

     430 alMetValGlnIleValProArgValTyrTyrHisProGluGluValVal 446
         ::|||||||||||||||||||||||||||||:::|||:::::::::::::
```

```
 1259 TTATGGTCCAAATTGTTCCCCGAGTATATTACTATCCCAAGAAAACAATT 1308

  447 LeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSe 463
      |||||||||||||||||||||   :::|||   |||:::::::|||:::||
 1309 CTCGATGAATATGATTACAGGAACCATCGACAAAAGAAAAAACCCATATC 1358

  463 rLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyT 480
      |||||||||||||||||||||||||||||||:::   :::|||||||   |
 1359 CCTGACACTCGCAGTAATGCTCGGACTCGGAGTGATAACAGGTGTGAGAA 1408

  480 hrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeu 496
      |||||||||||||||||:::|||||||||||||||||||||||   ||||||
 1409 CAGGAACTGCAGCTTTAGTTACAGGACCTCAGCAGCTAGAAACAGGACTT 1458

  497 GlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSe 513
      ::::::||||||      :::|||   :::|||:::||||||||||:::||
 1459 AGTAACCTACATCAAATTGTAACAGGAAATCTCCAAGCCCTAAAAAAATC 1508

  513 rValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuG 530
      |||||||||||||||||:::|||||||||||||||||||||||||||||||
 1509 TGTCAGTAACCTGGAAAAATCCCTAACCTCCTTATCTGAAGTAGTTCTAC 1558

  530 lnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCys 546
      |||||::::::|||||||||||||||||||::::::      ||||||
 1559 AGAATAAAAAAGGGTTAGATTTATTATTTCTAAAAAAAAGAAGATTATGT 1608

  547 AlaAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyAlaIl 563
      :::|||||||||||:::|||||||||||||||||||||||||||||||||
 1609 GTAGCCTTAAAGGAGAAATGCTGTTTTTATGTAGATCATTCAGGGGCCAT 1658

  563 eArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArgArgG 580
      |||||||||||||||||||||||:::|||:::|||||||:::||||||||||
 1659 CAGAGACTCCATGAACAAGCTTAAAAAAAGGTTGGAGAAACGTCGAAGGG 1708

  580 luArgGluAla.AspGlnGlyTrpPheGluGlyTrpPheAsnArgSerPr 596
      ||:::|||:::    |||   ||||||:::   ||||||||||||||||
 1709 AAAAGGAAACTTACTCAAAGATGGTTTAAAAGATGGTTCAACAGGTCTCC 1758

  596 oTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuL 613
      ||||::::::||||||||||:::||||||||||||||||:::||||||||
 1759 TTGGTTGGCTACCCTACTTTCTACTTTAACAGGACCCTTAATAGTCCTCC 1808

  613 euLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPhe 629
      |||||||||||||||||||||||||||:::||||||:::::::::||||||
 1809 TCCTGTTACTCACAGTTGGGCCATGTATTATTAACAAGTTAATTGCCTTC 1858

  630 ValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGlnTy 646
      :::|||:::|||:::||||||||||||||||||||||||||||||||||
 1859 ATTAGAAAACGAATAAGTGCAGTCCAGATCATGGTACTTAGACAACAGTA 1908

  646 rGln 647
      ||||
 1909 CCAA 1912

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-24

seq_documentation_block:
; Sequence 24, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
;  APPLICANT: Banerjee, Papia T.
;  APPLICANT: Patience, Clive
;  APPLICANT: Andersson, Goran K.
;  TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
;  TITLE OF INVENTION: Use
;  FILE REFERENCE: 61750-267
;  CURRENT APPLICATION NUMBER: US/09/376,781
;  CURRENT FILING DATE: 1999-08-18
;  EARLIER APPLICATION NUMBER: 60/097,015
;  EARLIER FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 24
;   LENGTH: 1493
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
;   OTHER INFORMATION: sequence of the 3' end of the PERV-D env region.
US-09-376-781-24


alignment_scores:
           Quality: 1675.00              Length:    489
             Ratio:  4.095                Gaps:      8
 Percent Similarity:  83.640   Percent Identity: 66.871

alignment_block:
 US-09-171-553B-10 x US-09-376-781-24   ..

 Align seg 1/1  to: US-09-376-781-24  from: 1  to: 1493

  163 ProGlyLysTyrLysMetMetLysLeu...TyrLysAspLysSer...Cy 177
      |||:::::::::::::      |||      |||:::      |||   ||
    4 CCCAATAACCACCGGACCTGGAAAACGTCATACAGGTACCATTCTGGGTG 53

  177 sSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyL 194
      |   |||||||||||||||||||||||||||||||||||||||||||||   |
   54 TTTTCCCTCAGACCTAGATTATCTTAAAATAAGTTTCACCGAAAAAAAAA 103

  194 ysGlnGluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyIleVal 210
      ||||||||||||||   |||||||||||||||||||||||||||||:::
  104 AACAAGAAAATATCCTAAAATGGATAAATGGTATGTCCTGGGGAATAATA 153

  211 PheTyrLys...TyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLe 226
      :::|||      :::             ||||||   ||||||||||||||
  154 TATTATACAGGTTGGGACAGACAACCAGGCTCCATTCTAACCATCCGACT 203

  226 uArgIleGluThrGlyThrGluProProValAlaMetGlyProAspLysV 243
      |:::|||   :::      |||||||||:::|||:::||||||:::   |
  204 TAAAATA...AGCCAGCTAGAGCCTCCAATGGCTATAGGACCGAATACGG 250

  243 alLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPro 259
      ||||||:::   |||      |||:::      |||      :::::
  251 TCTTAACGGGTCAAAGAACCCCAACCCCAGGACCATCCTCTGATATA... 297

  260 ValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerAs 276
                     :::   :::   |||   |||:::      |||||
  298 ..............ACTTCTAAATTAGACCCCACTGAG......TCTAA 326

  276 nSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGlyV 293
      ||||||||||
  327 CAGCACGACT..................................... 336

  293 alProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPhe 309
              |||||||||   :::|||||||||||||||||||||||||||||
  337 ........AAAACGGGGACAAAACTTTTTAGTCTCATCCAGGGAGCTTTT 378

  310 GlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCy 326
      ||||||:::|||||||||   |||:::|||||||||||||||||||||||
  379 CAAGCTCTTAACTCCACGACTCCAGAGGCTACCTCTTCTTGTTGGCTTTG 428

  326 sLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluArgLysP 343
      ||||:::|||||||||||||||||:::   |||||||||:::::||||
  429 CTTAACTTCGGGCCCACCTTACTATAAGAAAATGGCTAAAAGAGAAAAAT 478

  343 heAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAsn 359
      ||||||||||||||:::|||||||:::|||||||||||||||||||:::|||
  479 TCAATGTGACAAAAAAACATAGAGACCAATGTACATGGGGATCCCAAAAT 528

  360 LysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAl 376
      |||||||||||||||||||||||||||||   |||||||||   |||::
  529 AAGCTTACCCTTACTGAGGTTTCTGGAAAGGACACCTGCATAAAAAAGGT 578

  376 aProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGlnA 393
```

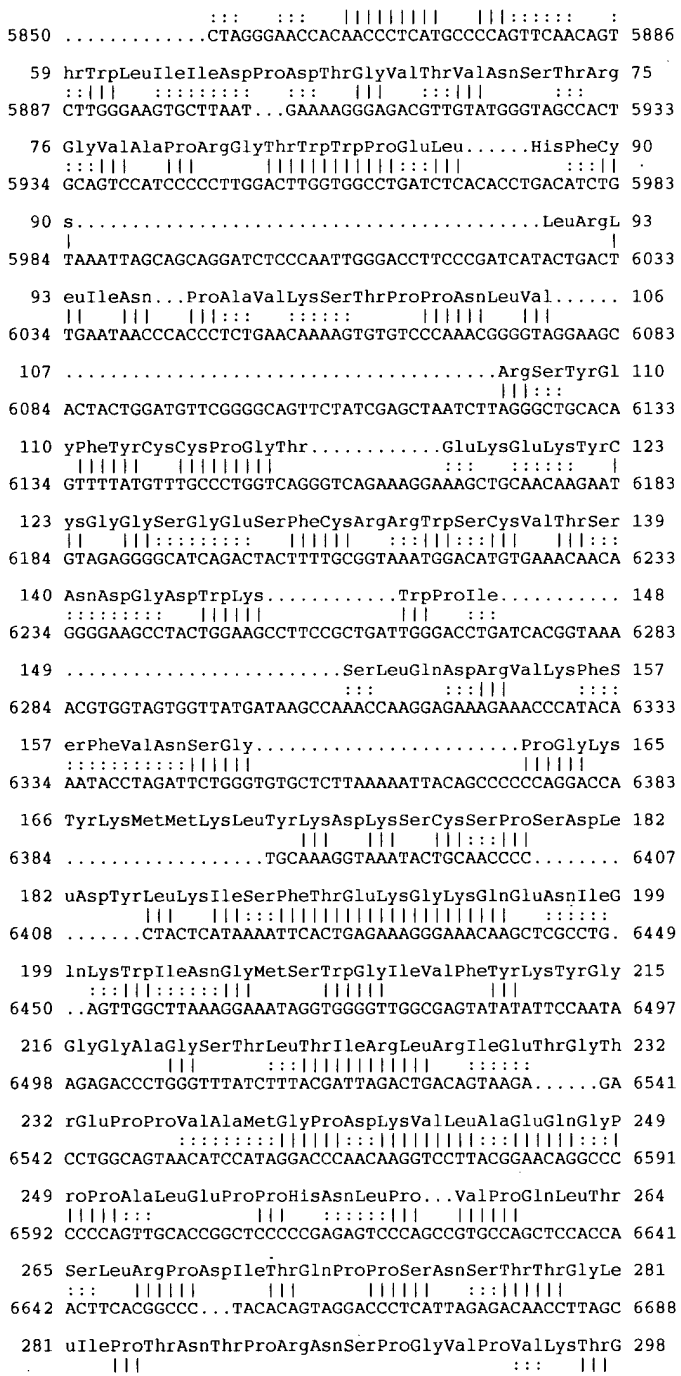
```
     :||||||||||||||||||||||         |||   :::::::::|||:
   579 TCCCCCATCCCACCAACACCTTTACAACCACACTGAAGCCTTTAATCAAA 628

   393 laSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCys 409
     ::|||||:::|||||||||||||||||||||:::|||||||||||||||
   629 CCTCTGAGAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGT 678

   410 AsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSerLy 426
     ||||||||||||||||||||||||||||   ||||||||||||:::||
   679 AATACTGGATTAACCCCTTGTGTTTCCACCTTGGTTTTCAACCAAACTAA 728

   426 sAspLeuCysValMetValGlnIleValProArgValTyrTyrHisProG 443
     ||||:::   :::|||||||||||||||||||||||||||||:::|||:
   729 AGACTTTTACATTATGGTCCAAATTGTTCCCCGAGTATATTACTATCCCA 778

   443 luGluValValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArg 459
     ::::::::::|||||||||||||||||||||   :::|||   |||:::
   779 AGAAAACAATTCTCGATGAATATGATTACAGGAACCATCGACAAAAGAAA 828

   460 GluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaVa 476
     :::|||:::|||||||||||||||||||||||||||||:::   ::
   829 AAACCCATATCCCTGACACTCGCAGTAATGCTCGGACTCGGAGTGATAAC 878

   476 lGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuG 493
     :||||||   ||||||||||||||||||:::||||||||||||||||||||
   879 AGGTGTGAGAACAGGAACTGCAGCTTTAGTTACAGGACCTCAGCAGCTAG 928

   493 luLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAla 509
     ||   ||||||:::::||||||      :::|||   :::|||:::|||
   929 AAACAGGACTTAGTAACCTACATCAAATTGTAACAGGAAATCTCCAAGCC 978

   510 LeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGl 526
     ||||||:::|||||||||||||||||:::|||||||||||||||||||||
   979 CTAAAAAAATCTGTCAGTAACCTGGAAAAATCCCTAACCTCCTTATCTGA 1028

   526 uValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluG 543
     ||||||||||||||:::::|||||||||||||||||||||:::::::
  1029 AGTAGTTCTACAGAATAAAAAAGGGTTAGATTTATTATTTCTAAAAAAAA 1078

   543 lyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHis 559
        ||||||:::|||||||||||||:::||||||||||||||||||||
  1079 GAAGATTATGTGTAGCCTTAAAGGAGAAATGCTGTTTTTATGTAGATCAT 1128

   560 SerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGluAr 576
     ||||||||||||||||||||||||||||||:::|||:::||||||::
  1129 TCAGGGGCCATCAGAGACTCCATGAACAAGCTTAAAAAAAGGTTGGAGAA 1178

   576 gArgArgArgGluArgGluAla.AspGlnGlyTrpPheGluGlyTrpPhe 592
     :|||||||||||:::|||:::    |||   |||||||:::   ||||||
  1179 ACGTCGAAGGGAAAAGGAAACTTACTCAAAGATGGTTTAAAAGATGGTTC 1228

   593 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLe 609
     ||||||||||||||:::::|||||||||||:::|||||||||||||||
  1229 AACAGGTCTCCTTGGTTGGCTACCCTACTTTCTACTTTAACAGGACCCTT 1278

   609 uValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgP 626
     |:::|||||||||||||||||||||||||||||||:::|||||||::::
  1279 AATAGTCCTCCTCCTGTTACTCACAGTTGGGCCATGTATTATTAACAAGT 1328

   626 heValAlaPheValArgGluArgValSerAlaValGlnIleMetValLeu 642
     :::::||||||:::|||:::|||:::||||||||||||||||||||||
  1329 TAATTGCCTTCATTAGAAAACGAATAAGTGCAGTCCAGATCATGGTACTT 1378

   643 ArgGlnGlnTyrGln 647
     |||||||||||||||
  1379 AGACAACAGTACCAA 1393

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-7

seq_documentation_block:
; Sequence 7, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 6312
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4058)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4059)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4060)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4061)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4246)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4247)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4248)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4249)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-7


alignment_scores:
            Quality: 1544.50            Length:    660
              Ratio:  3.300              Gaps:     15
 Percent Similarity:  70.909   Percent Identity: 46.970

alignment_block:
 US-09-171-553B-10 x US-09-011-745-7   ..

 Align seg 1/1  to: US-09-011-745-7  from: 1  to: 6312

     47 IleAspSerSerAsnProHisArgProLeuSerProThrTrpLeuIleIl 63
        ::::::::::::|||||||||:::|||::::::   ||||||   :::::
   1675 CTGCAAAATAAGAACCCCCACCAGCCCATGACCCTCACTTGGCAGGTACT 1724

     63 eAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProA 80
        ::::   :::||||||   :::|||   :::|||::::::|||   |||
   1725 GTCC...CAAACTGGAGACGTTGTCTGGGATACAAAGGCAGTCCAGCCCC 1771
```

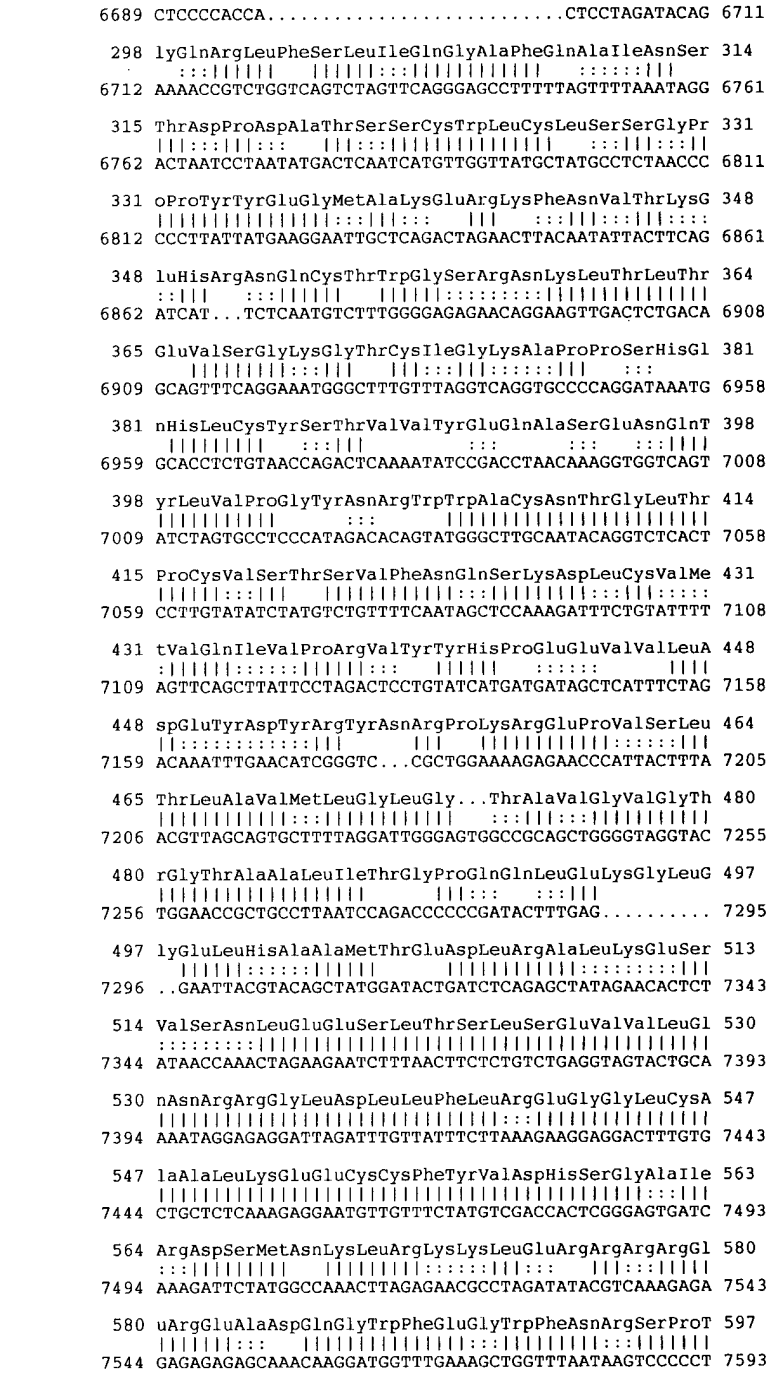
80 rgGlyThrTrpTrpProGluLeu.....HisPheCysLeuArgLeu... 93
1772 CTTGGACTTGGTGGCCCACACTTAAACCTGATGTATGTGCCTTGGCGGCT 1821
94 .....IleAsnProAlaValLy 99
1822 AGTCTTGAGTCCTGGGATATCCCGGGAACCGATGTCTCGTCCTCTAAACG 1871
99 sSerThrProProAsn..... 104
1872 AGTCAGACCTCCGGACTCAGACTATACTGCCGCTTATAAGCAAATCACCT 1921
105 .....LeuValArgSer 108
1922 GGGGAGCCATAGGGTGCAGCTACCCTCGGGCTAGGACTAGAATGGCAAGC 1971
109 TyrGlyPheTyrCysCysPro.....GlyThrGluLysGluLy 121
1972 TCTACCTTCTACGTATGTCCCCGGGATGGCCGGACCCTTTCAGAAGCTAG 2021
121 sTyrCysGlyGlySerGlyGluSerPheCysArgArgTrpSerCysValT 138
2022 AAGGTGCGGGGGGCTAGAATCCCTATACTGTAAAGAATGGGATTGTGAGA 2071
138 hrSerAsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgVal 154
2072 CCACGGGGACCGGTTATTGG...CTATCTAAATCCTCAAAAGACCTCATA 2118
155 LysPheSerPheValAsnSerGlyProGlyLysTyrLysMetMetLysLe 171
2119 ACTGTAAAATGGGACCAAAATAGCGAATGGACTCAAAAATTTCAACAGTG 2168
171 uTyrLysAspLysSerCysSerProSerAspLeuAspTyrLeuLysIleS 188
2169 TCACCAGACCGGCTGGTGTAACCCC.....CTTAAAATAG 2203
188 erPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrpIleAsnGly 204
2204 ATTTCACAGACAAAGGAAAATTATCC.....AAGGACTGGATAACGGGA 2247
205 MetSerTrpGlyIleValPheTyrLysTyrGlyGlyGlyAlaGlySerTh 221
2248 AAAACCTGGGGATTAAGATTCTAT...GTGTCTGGACATCCAGGCGTACA 2294
221 rLeuThrIleArgLeuArgIleGluThrGlyThrGluProProValAlaM 238
2295 GTTCACCATTCGCTTAAAAATCACC.....AACATGCCAGCTGTGGCAG 2338
238 etGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeuGlu... 253
2339 TAGGTCCTGACCTCGTCCTTGTGGAACAAGGACCTCCTAGAACGTCCCTC 2388
254 .....ProProHisAsnLeuProValProGlnLeuTh 264
2389 GCTCTCCCACCTCCTCTTCCCCCAAGGGAAGCGCCACCGCCATCTCTC.. 2436
264 rSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrGlyL 281
2437 .....CCCGACTCTAACTCCACAGCCCTGG 2461
281 euIleProThrAsnThrProArg.....Asn 289
2462 CGACTAGTGCACAAACTCCCACGGTGAGAAAAACAATTGTTACCCTAAAC 2511
290 SerProGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGl 306
2512 ACTCCG.....CCTCCCACCACAGGCGACAGACTTTTTGATCTTGTGCA 2555
306 nGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerC 323
2556 GGGGGCCTTCCTAACCTTAAATGCTACCAACCCAGGGGCCACTGAGTCTT 2605
323 ysTrpLeuCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLys 339
2606 GCTGGCTTTGTTTGGCCATGGGCCCCCCTTATTATGAAGCAATAGCCTCA 2655
340 GluArgLysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGl 356
2656 TCAGGAGAGGTCGCCTACTCCACCGAC...CTTGACCGGTGCCGCTGGGG 2702
356 ySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysI 373
2703 GACCCAAGGAAAGCTCACCCTCACTGAGGTCTCAGGACACGGGTTGTGCA 2752
373 leGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValVal 389
2753 TAGGAAAGGTGCCCTTTACCCATCAGCATCTCTGCAATCAGACCCTATCC 2802
390 TyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTr 406
2803 ATCAATTCCTCCGGAGACCATCAGTATCTGCTCCCCTCCAACCATAGCTG 2852
406 pTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheA 423
2853 GTGGGCTTGCAGCACTGGCCTCACCCCTTGCCTCTCCACCTCAGTTTTTA 2902
423 snGlnSerLysAspLeuCysValMetValGlnIleValProArgValTyr 439
2903 ATCAGACTAGAGATTTCTGTATCCAGGTCCAGCTGATTCCTCGCATCTAT 2952
440 TyrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnAr 456
2953 TACTATCCTGAAGAAGTTTTGTTACAGGCCTATGACAATTCTCACCCCAG 3002
456 gProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuG 473
3003 GACTAAAAGAGAGGCTGTCTCACTTACCCTAGCTGTTTTACTGGGGTTGG 3052
473 lyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyPro 489
3053 GAATCACGGCGGGAATAGGTACTGGTTCAACTGCCTTAATTAAAGGACCT 3102
490 GlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAs 506
3103 ATAGACCTCCAGCAAGGCCTGACAAGCCTCCAGATCGCCATAGATGCTGA 3152
506 pLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrS 523
3153 CCTCCGGGCCCTCCAAGACTCAGTCAGCAAGTTAGAGGACTCACTGACTT 3202
523 erLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPhe 539
3203 CCCTGTCCGAGGTAGTGCTCCAAAATAGGAGAGGCCTTGACTTGCTGTTT 3252
540 LeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTy 556
3253 CTAAAAGAAGGTGGCCTCTGTGCGGCCCTAAAGGAAGAGTGCTGTTTTTA 3302
556 rValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysL 573
3303 CATAGACCACTCAGGTGCAGTACGGGACTCCATGAAAAAACTCAAAGAAA 3352
573 ysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGlu 589
3353 AACTGGATAAAAGACAGTTAGAGCGCCAGAAAAGCCAAAACTGGTATGAA 3402
590 GlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuTh 606
3403 GGATGGTTCAATAACTCCCCTTGGTTCACTACCCTGCTATCAACCATCGC 3452
606 rGlyProLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuI 623
3453 TGGGCCCCTATTACTCCTCCTTCTGTTGCTCATCCTCGGGCCATGCATCA 3502
623 leAsnArgPheValAlaPheValArgGluArgValSerAlaValGlnIle 639

```
        |||||||||:::|||   ||||||:::::::|||:::|||:::||||||
    3503 TCAATCGATTAGTTCAATTTGTTAAAGACAGGATCTCAGTAGTCCAGGCT 3552

     640 MetValLeuArgGlnGlnTyrGlnGlyLeu 649
         :::||||||   ||||||||||:::   |||
    3553 TTAGTCCTGACTCAACAATACCACCAGCTA 3582

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-1

seq_documentation_block:
; Sequence 1, Application US/08716351A
; Patent No. 6033905
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Gibbon-Ape Leukemia Virus-Based
;    TITLE OF INVENTION:  Retroviral Vectors
;    NUMBER OF SEQUENCES:  5
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/716,351A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/03784
;      FILING DATE:  06-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian,, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  15280-128-1PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..8535
;      OTHER INFORMATION:  /standard_name= "GaLV SEATO Genome"
US-08-716-351A-1


alignment_scores:
            Quality: 1522.50            Length:    656
              Ratio:  3.267              Gaps:     16
 Percent Similarity:  71.037   Percent Identity: 46.799

alignment_block:
 US-09-171-553B-10 x US-08-716-351A-1   ..

 Align seg 1/1  to: US-08-716-351A-1  from: 1  to: 8535

      47 IleAspSerSerAsnProHisArgProLeuSerProThrTrpLeuIleIl 63
         ::::::::::::|||||||||:::|||:::::::   ||||||   :::::
    6051 CTGCAAAATAAGAACCCCCACCAGCCCATGACCCTCACTTGGCAGGTACT 6100

      63 eAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProA 80
         ::::   :::||||||   :::|||   :::|||:::::::|||   |||
    6101 GTCC...CAAACTGGAGACGTTGTCTGGGATACAAAGGCAGTCCAGCCCC 6147

      80 rgGlyThrTrpTrpProGluLeu......HisPheCysLeuArgLeu... 93
            ||||||||||||   |||           |||
    6148 CTTGGACTTGGTGGCCCACACTTAAACCTGATGTATGTGCCTTGGCGGCT 6197

      94 ..............................IleAsnProAlaValLy 99
                                       ::::::   :::   ::
    6198 AGTCTTGAGTCCTGGGATATCCCGGGAACCGATGTCTCGTCCTCTAAACG 6247

      99 sSerThrProProAsn................................ 104
         :      ||||||:::
    6248 AGTCAGACCTCCGGACTCAGACTATACTGCCGCTTATAAGCAAATCACCT 6297

     105 .....................................LeuValArgSer 108
                                              :::   |||
    6298 GGGGAGCCATAGGGTGCAGCTACCCTCGGGCTAGGACTAGAATGGCAAGC 6347

     109 TyrGlyPheTyrCysCysPro............GlyThrGluLysGluLy 121
               ||||||   ||||||                ::::::   ::
    6348 TCTACCTTCTACGTATGTCCCCGGGATGGCCGGACCCTTTCAGAAGCTAG 6397

     121 sTyrCysGlyGlySerGlyGluSerPheCysArgArgTrpSerCysValT 138
         :   |||||||||      :::   :::|||::::::|||:::|||   |
    6398 AAGGTGCGGGGGGCTAGAATCCCTATACTGTAAAGAATGGGATTGTGAGA 6447

     138 hrSerAsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgVal 154
         ||::::::   |||  |||           |||   :::|||   :::
    6448 CCACGGGGACCGGTTATTGG...CTATCTAAATCCTCAAAAGACCTCATA 6494

     155 LysPheSerPheValAsnSerGlyProGlyLysTyrLysMetMetLysLe 171
               ::::::   :::::::::           |||:::::::::
    6495 ACTGTAAAATGGGACCAAAATAGCGAATGGACTCAAAAATTTCAACAGTG 6544

     171 uTyrLysAspLysSerCysSerProSerAspLeuAspTyrLeuLysIleS 188
          ::::::         |||:::|||               |||||||||:
    6545 TCACCAGACCGGCTGGTGTAACCCC..............CTTAAAATAG 6579

     188 erPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrpIleAsnGly 204
         ::|||||:::|||||||||   :::       :::   ||||||:::|||
    6580 ATTTCACAGACAAAGGAAAATTATCC......AAGGACTGGATAACGGGA 6623

     205 MetSerTrpGlyIleValPheTyrLysTyrGlyGlyGlyAlaGlySerTh 221
            :::|||||:::   ||||||      :::|||      |||
    6624 AAAACCTGGGGATTAAGATTCTAT...GTGTCTGGACATCCAGGCGTACA 6670

     221 rLeuThrIleArgLeuArgIleGluThrGlyThrGluProProValAlaM 238
          :::|||||||||||||:::|||         :::   |||   ||||||:
    6671 GTTCACCATTCGCTTAAAAATCACC......AACATGCCAGCTGTGGCAG 6714

     238 etGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeuGlu... 253
         ::|||||||||   ||||||:::||||||||||||||||      :::
    6715 TAGGTCCTGACCTCGTCCTTGTGGAACAAGGACCTCCTAGAACGTCCCTC 6764

     254 ..................ProProHisAsnLeuProValProGlnLeuTh 264
                             ||||||::::::   |||   |||:::|||
    6765 GCTCTCCCACCTCCTCTTCCCCCAAGGGAAGCGCCACCGCCATCTCTC.. 6812

     264 rSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrGlyL 281
                                  |||   ||||||||||||:::
    6813 .........................CCCGACTCTAACTCCACAGCCCTGG 6837

     281 euIleProThrAsnThrProArg.......................Asn 289
                ::::::||||||                           |||
    6838 CGACTAGTGCACAAACTCCCACGGTGAGAAAAACAATTGTTACCCTAAAC 6887

     290 SerProGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGl 306
         :::|||      |||      ||||||:::|||||||||:::|||:::||
    6888 ACTCCG......CCTCCCACCACAGGCGACAGACTTTTTGATCTTGTGCA 6931

     306 nGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerC 323
         ||||||||||   :::::::|||:::|||:::|||   ||||||:::||||
    6932 GGGGGCCTTCCTAACCTTAAATGCTACCAACCCAGGGGCCACTGAGTCTT 6981

     323 ysTrpLeuCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLys 339
         ||||||||||||||:::   ||||||||||||||||||:::::::|||:::
    6982 GCTGGCTTTGTTTGGCCATGGGCCCCCCTTATTATGAAGCAATAGCCTCA 7031

     340 GluArgLysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGl 356
```

```
     :::   :::         :::   :::      :::::::|||   |||||
7032 TCAGGAGAGGTCGCCTACTCCACCGAC...CTTGACCGGTGCCGCTGGGG 7078

 356 ySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysI 373
     |:::::::::|||||||||||||||||||||||||||||   |||   ||||
7079 GACCCAAGGAAAGCTCACCCTCACTGAGGTCTCAGGACACGGGTTGTGCA 7128

 373 leGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValVal 389
     ||||||||:::|||   :::|||||||||||||||   :::|||:::
7129 TAGGAAAGGTGCCCTTTACCCATCAGCATCTCTGCAATCAGACCCTATCC 7178

 390 TyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTr 406
     :::::::::::::::::::|||||||||:::|||:::   :::   ||
7179 ATCAATTCCTCCGGAGACCATCAGTATCTGCTCCCCTCCAACCATAGCTG 7228

 406 pTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheA 423
     ||||||||||:::|||||||||||||||||||:::|||||||||||||||
7229 GTGGGCTTGCAGCACTGGCCTCACCCCTTGCCTCTCCACCTCAGTTTTTA 7278

 423 snGlnSerLysAspLeuCysValMetValGlnIleValProArgValTyr 439
     |||||::::::|||:::|||::::::||||||::::::||||||:::|||
7279 ATCAGACTAGAGATTTCTGTATCCAGGTCCAGCTGATTCCTCGCATCTAT 7328

 440 TyrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnAr 456
     |||:::|||||||||||||:::|||:::   ||||||      :::   ||
7329 TACTATCCTGAAGAAGTTTTGTTACAGGCCTATGACAATTCTCACCCCAG 7378

 456 gProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuG 473
     |   |||||||||   ||||||||||||||||||||:::||||||||||
7379 GACTAAAAGAGAGGCTGTCTCACTTACCCTAGCTGTTTTACTGGGGTTGG 7428

 473 lyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyPro 489
     ||   ::::::|||:::|||||||||::::::|||||||||   ||||||
7429 GAATCACGGCGGGAATAGGTACTGGTTCAACTGCCTTAATTAAAGGACCT 7478

 490 GlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAs 506
     :::|||::::::||||||   :::|||:::   |||:::      ||
7479 ATAGACCTCCAGCAAGGCCTGACAAGCCTCCAGATCGCCATAGATGCTGA 7528

 506 pLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrS 523
     |||||||||||||:::::|||||||||:::||||||:::||||||||||
7529 CCTCCGGGCCCTCCAAGACTCAGTCAGCAAGTTAGAGGACTCACTGACTT 7578

 523 erLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPhe 539
     ||||||||||||||||||||||||||||||||||||||||||||||||||
7579 CCCTGTCCGAGGTAGTGCTCCAAAATAGGAGAGGCCTTGACTTGCTGTTT 7628

 540 LeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTy 556
     |||:::||||||||||||||||||||||||||||||||||||||||||||
7629 CTAAAAGAAGGTGGCCTCTGTGCGGCCCTAAAGGAAGAGTGCTGTTTTTA 7678

 556 rValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysL 573
     |:::|||||||||||||||:::|||||||||||:::|||||:::::::|
7679 CATAGACCACTCAGGTGCAGTACGGGACTCCATGAAAAAACTCAAAGAAA 7728

 573 ysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGlu 589
     |||||:::::|||:::   ||||||:::   :::|||:::|||:::|||
7729 AACTGGATAAAAGACAGTTAGAGCGCCAGAAAAGCCAAAACTGGTATGAA 7778

 590 GlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuTh 606
     ||||||||||:::|||||||||::|||||||||||||||:::::::::
7779 GGATGGTTCAATAACTCCCCTTGGTTCACTACCCTGCTATCAACCATCGC 7828

 606 rGlyProLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuI 623
     :|||||||||:::::|||||||||||||||   :::|||||||||:::|
7829 TGGGCCCCTATTACTCCTCCTTCTGTTGCTCATCCTCGGGCCATGCATCA 7878

 623 leAsnArgPheValAlaPheValArgGluArgValSerAla.ValGlnIl 639
     |||||::::::|||   |||:::::::::::|||:::|||||| |||:::||
7879 TCAATAAGTTAGTTCAATTCATCAATGATAGGATAAGTGCATGTTAAAAT 7928

 639 eMetValLeuArgGln 644
     |:::|||||||||||||
```

```
    7929 TCTGGTCCTTAGACAA 7944

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-075-272-1

seq_documentation_block:
; Sequence 1, Application US/09075272
; Patent No. 6136598
;  GENERAL INFORMATION:
;    APPLICANT:  MILLER, A. DUSTY
;    APPLICANT:  WOLGAMOT, GREG
;    APPLICANT:  BONHAM, LYNN
;    TITLE OF INVENTION:  MUS DUNNI ENDOGENOUS RETROVIRAL
;    TITLE OF INVENTION:  PACKAGING CELL LINES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,272
;      FILING DATE:  08-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/046,140
;      FILING DATE:  09-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Poor, Brian W.
;      REGISTRATION NUMBER:  32,928
;      REFERENCE/DOCKET NUMBER:  14538A-003710
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-075-272-1


alignment_scores:
            Quality: 1478.50              Length:    728
              Ratio:  3.080                Gaps:     24
 Percent Similarity:  65.934   Percent Identity: 44.368

alignment_block:
 US-09-171-553B-10 x US-09-075-272-1   ..

 Align seg 1/1  to: US-09-075-272-1  from: 1  to: 8655

       6 SerTrpArgHis.LeuProThrArgGlyGlyGluProLysArgLeuArgI 22
         ||||||   ||| |||   |||      |||   |||   :::::::::
    5711 TCCTGGATCCACGCTTCACACGTCAAGAGGGCCCCCAGTCAAGATGAAGA 5760

      22 lePro.....................................LeuSer 25
           |||                                     ||||||
    5761 AACCCACGAAGACAACTGGGCTGTGGAAGCCACTGATAACCCTCTTAAGC 5810

      26 PheAlaSerIleAlaTrpPheLeuThrLeuThrIleThrProGlnAlaSe 42
         ||||||   :::|||         :::::|||:::
    5811 TTCGCTTGCGTCGCAGGAGCCCCCTCCATCACCCTGGAC..... 5849

      42 rSerLysArgLeuIleAspSerSerAsnProHisArgProLeuSerProT 59
```





```
      597 rpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeu 613
          ||:::|||||||||||||||||:::::::::|||||||||:::::::||||||
     7594 GGCTCACCACTCTCCTTTCCACTATAGCGGGACCCTTGATTATACTTCTG 7643

      614 LeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheVa 630
          ||||||||||||   |||||||||:::::::|||:::::::|||||||||::
     7644 CTTTTGCTTACTTTTGGCCCCTGCATTCTTAATAAGTTAGTAGCCTTTAT 7693

      630 lArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrG 647
          :|||||||||:::::||||||||||:::|||||||||:::||||||||||
     7694 TAGAGAAAGAATAAATGCAGTACAGGTTATGGTATTAAAACAACAATATC 7743

      647 lnGlyLeuLeuSerGlnGlyGluThrAspLeu 657
          ||   ::::::::::::::::|||:::::|||
     7744 AG...GTCTTCCAGGAGGCCGAAAACTCGCTC 7772

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-300A-8

seq_documentation_block:
; Sequence 8, Application US/08110300A
; Patent No. 5643756
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/110,300A
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-110-300A-8


alignment_scores:
            Quality: 1329.00              Length:    696
              Ratio:  2.973                Gaps:     23
 Percent Similarity:  64.224   Percent Identity: 42.816

alignment_block:
 US-09-171-553B-10 x US-08-110-300A-8   ..

 Align seg 1/1  to: US-08-110-300A-8  from: 1  to: 8323

      16 GluProLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPh 32
         :::|||:::   |||   |||||||||   :::                ||
   5817 GACCCGCGGGACCTCCTAATCCCCTTAATTCTC.............TT 5851

      32 eLeuThrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspS 49
         ||||:::|||                  :::::::::           :
   5852 CCTGTCTCTC.............AAAGGGGCCAGATCCGCAGCACCCG 5886

      49 erSerAsnProHisArgProLeuSerProThrTrpLeuIleIleAspPro 65
         ::|||:::||||||:::      :::   ||||||   :::   :::
   5887 GCTCCAGCCCTCACCAGGTCTACAACATTACCTGGGAAGTGACCAATGGG 5936

      66 AspThrGlyValThrValAsnSerThrArgGlyValAlaProArgGlyTh 82
         |||         ||||||   :::      |||      |||       ||
   5937 GATCGG...GAGACAGTATGGGCAATATCAGGCAACCACCCTCTGTGGAC 5983

      82 rTrpTrpProGluLeu..................................... 87
         ||||||||||   |||
   5984 TTGGTGGCCAGTCCTCACCCCAGATTTGTGTATGTTAGCTCTCAGTGGGC 6033

      88 .....HisPhe..................................... 89
              |||:::
   6034 CGCCCCACTGGGGGCTAGAGTATCAGGCCCCCTATTCCTCGCCCCCGGGG 6083

      90 .....................................CysLeuArgLe 93
                                             |||   |||
   6084 CCCCCTTGTTGCTCAGGGAGCAGCGGGAACGTTGCAGGCTGTGCCAGAGA 6133

      93 uIleAsnProAlaValLysSerThrProPro.................A 104
            |||       :::   |||       |||                |
   6134 CTGCAACGAGCCCTTGACCTCCCTCACCCCTCGGTGCAACACTGCCTGGA 6183

     104 snLeuValArg.......................SerTyrGlyPheTyr 112
         ||   ::::::                      |||   |||||||||
   6184 ACAGACTTAAGCTGGACCAGGTAACTCATAAATCAAGTGAGGGATTTTAT 6233

     113 CysCysProGlyThrGluLys.........GluLysTyrCysGlyGlySe 126
            |||||||||:::::::::           |||   |||||||||
   6234 GTCTGCCCCGGGTCACATCGCCCCCGGGAAGCCAAGTCCTGTGGGGGTCC 6283

     126 rGlyGluSerPheCysArgArgTrpSerCysValThrSerAsnAspGlyA 143
            :::   :::|||      |||:::|||   |||::::::
   6284 AGACTCCTTCTACTGTGCCTCTTGGGGCTGCGAGACAACCGGTAGAGTAT 6333

     143 spTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159
         ||||||   |||   |||       |||   :::            |||
   6334 ACTGGAAG...CCCTCCTCTTCTTGGGACTACATCACA.........GTA 6371

     160 AsnSerGlyProGlyLysTyrLysMetMetLysLeuTyrLysAsp...Ly 175
         :::::::::      :::   :::   :::::::::   ||||||   ||
   6372 GACAACAATCTCACCTCTAACCAGGCTGTTCAGGTATGCAAAGACAATAA 6421

     175 sSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluL 192
         |   |||:::|||               |||   |||   ||||||:::
   6422 GTGGTGCAATCCC.............TTGGCTATCCGGTTTACAAACG 6456

     192 ysGlyLysGlnGluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGly 208
         |||||||||      :::   :::|||   :::|||      ||||||
   6457 CCGGGAAACAG......GTCACCTCATGGACAACTGGACACTATTGGGGT 6500

     209 IleValPheTyrLysTyrGlyGlyGlyAlaGlySerThrLeuThrIleAr 225
         :::   :::|||      |||         |||   |||:::   |||||
   6501 CTACGTCTTTATGTCTCTGGACAGGACCCAGGGCTTACTTTCGGGATCCG 6550

     225 gLeuArgIleGluThrGlyThrGluProProValAlaMetGlyProAspL 242
         ||||      :::   :::      |||   :::   :::||||||:::
   6551 ACTCAGTTATCAA...AATCTAGGACCTCGGATCCCAATAGGACCAAACC 6597

     242 ysValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeu 258
         |||||||||:::|||           |||:::   |||
   6598 CCGTCCTGGCAGACCAA...........CTTTCGTTCCCGCTACCTAAT 6635

     259 ProValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSe 275
         |||:::|||               :::|||      :::   |||   ||
```

```
6636 CCCCTACCC..............AAACCTGCCAAGTCTCCCCCCGCCTC 6670

 275 rAsnSerThrThrGlyLeuIle...ProThrAsnThrProArgAsnSerP 291
     |:::||||||      ||||||   |||:::   ||||||   :::   |
6671 TAGTTCGACTCCCACATTGATTTCCCCGTCCCCCACTCCCACTCAGCCCC 6720

 291 roGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGly 307
     ||      |||:::   ||||||:::||||||:::::::|||:::||||||
6721 CG......CCAGCAGGAACGGGAGACAGATTACTAAATCTAGTACAGGGA 6764

 308 AlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTr 324
     |||:::||||||:::|||   |||:::||||||   |||::::::|||||
6765 GCTTACCAGGCACTCAACCTTACCAACCCTGATAAAACTCAAGAGTGCTG 6814

 324 pLeuCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluA 341
     |||||||||   ||||||||||||||||||||||:::|||
6815 GTTATGCCTAGTGTCTGGACCCCCCTATTACGAGGGGGTTGCCGTCCTAG 6864

 341 rgLysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySer 357
         ::::::        :::         :::|||:::   :::|||
6865 GTACTTATTCCAACCATACCTCTGCCCCAGCTAACTGCTCCGTGGCCTCC 6914

 358 ArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGl 374
     ::::::||||||||||||:::||||||:::|||:::|||   ||||||||
6915 CAACACAAGCTGACCCTGTCCGAAGTGACTGGACGGGGACTCTGCATAGG 6964

 374 yLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrG 391
     |   :::|||   :::||||||   ||||||   :::|||:::      :
6965 AACAGTCCCAAAAACTCACCAGGCCCTGTGCAACACTACCCTT......A 7008

 391 luGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrp 407
     ::   ::::::   :::   |||||||||         :::       |||
7009 AGGCAGGCAAAGGGTCTTACTATCTAGTTGCCCCCACAGGAACTATGTGG 7058

 408 AlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGl 424
     ||||||||||||||||||||||||||:::|||::::::|||:::|||::
7059 GCATGTAACACTGGACTCACTCCATGCCTATCTGCCACCGTGCTTAATCG 7108

 424 nSerLysAspLeuCysValMetValGlnIleValProArgValTyrTyrH 441
     ::::   |||   |||||||:::|||::::::   |||||||||   ||||
7109 CACCACTGACTATTGCGTTCTCGTGGAATTATGGCCCAGGGTCACCTACC 7158

 441 isProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgPro 457
     |||||   :::   |||   ::::::::::::   :::   :::|||
7159 ATCCTCCCAGTTACGTCTATAGCCAGTTTGAA...AAATCCCATAGACAT 7205

 458 LysArgGluProValSerLeuThrLeuAlaValMetLeuGly..... 471
     |||||||||||||||||||||||||||||:::::::||||||
7206 AAAAGAGAACCAGTGTCCTTAACCTTGGCCTTATTATTAGGTGGGCTAAC 7255

 472 ....LeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleT 487
         |||   |||:::||||||||||||||||||:::|||||||:::
7256 TATGGGTGGCATCGCCGCGGGAGTAGGGACAGGAACTACCGCCCTGGTC. 7304

 487 hrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMet 503
       :::   ||||||::::::            :::||||||||||||:::
7305 ..GCCACCCAGCAGTTTCAG............CAGCTCCATGCTGCCGTA 7340

 504 ThrGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSe 520
       :::||||||:::   :::::::::|||::::::|||||||||:::||
7341 CAAGATGATCTCAAAGAAGTCGAAAAGTCAATTACTAACCTAGAAAAGTC 7390

 520 rLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspL 537
     ||||||||||||||||||||||||||||||||||||||||||||||||||
7391 TCTTACTTCGTTGTCTGAGGTTGTACTGCAGAATCGACGAGGCCTAGACC 7440

 537 euLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCys 553
     ||||||||||::||||||||||||||||||||||||||||||||||||||
7441 TGTTGTTCCTAAAAGAGGGAGGACTGTGTGCTGCCCTAAAAGAAGAATGT 7490

 554 CysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLe 570
     |||||||||:::||||||:::|||   :::||||||||||||   |||||
7491 TGTTTCTATGCTGACCATACAGGCCTAGTAAGAGATAGTATGGCCAAATT 7540
```

```
     570 uArgLysLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyT 587
         |||:::::::|||::::::|||::::::       |||::::::|||||||
    7541 AAGAGAGAGACTCTCTCAGAGACAAAAACTATTTGAGTCGAGCCAAGGAT 7590

     587 rpPheGluGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSer 603
         ||||||||||||||||||||||||||||||:::|||||||||:::|||
    7591 GGTTCGAAGGATGGTTTAACAGATCCCCCTGGTTTACCACGTTGATATCC 7640

     604 AlaLeuThrGlyProLeuValValLeuLeuLeuLeuThrValGlyPr 620
         ::::::   |||||||||::::::|||||||||:::|||      |||||
    7641 ACCATCATGGGGCCTCTCATTATACTCCTACTAATTCTGCTTTTTGGACC 7690

     620 oCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaV 637
         |||:::::::|||||::|||   ||||||:::::|||:::|||:::|
    7691 CTGCATTCTTAATCGATTAGTTCAATTTGTTAAAGACAGGATCTCAGTAG 7740

     637 alGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 649
         |||||   :::||||||   |||||||||:::   |||
    7741 TCCAGGCTTTAGTCCTGACTCAACAATACCACCAGCTA 7778
```

```
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-8

seq_documentation_block:
; Sequence 8, Application US/08886642
; Patent No. 5952474
;  GENERAL INFORMATION:
;    APPLICANT:  Pinter, Abraham
;    APPLICANT:  Kayman, Samuel
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  45 Rockefeller Plaza, Suite 2800
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/886,642
;      FILING DATE:  01-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/110,300
;      FILING DATE:  20-AUG-1993
;      APPLICATION NUMBER:  07/938,100
;      FILING DATE:  28-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hone, William J.
;      REGISTRATION NUMBER:  26,739
;      REFERENCE/DOCKET NUMBER:  07763/010002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212/765-5070
;      TELEFAX:  212/258-2291
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-886-642-8


alignment_scores:
            Quality: 1329.00              Length:    696
              Ratio:  2.973                Gaps:     23
 Percent Similarity:  64.224   Percent Identity: 42.816
```

```
alignment_block:
 US-09-171-553B-10 x US-08-886-642-8   ..

 Align seg 1/1  to: US-08-886-642-8  from: 1  to: 8323

      16 GluProLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPh 32
         :::|||:::   |||   |||||||||   :::                ||
    5817 GACCCGCGGGACCTCCTAATCCCCTTAATTCTC.............TT 5851

      32 eLeuThrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspS 49
         ||||:::|||                 :::::::::              :
    5852 CCTGTCTCTC.............AAAGGGGCCAGATCCGCAGCACCCG 5886

      49 erSerAsnProHisArgProLeuSerProThrTrpLeuIleIleAspPro 65
         ::|||:::|||||||:::      :::   ||||||   :::   :::
    5887 GCTCCAGCCCTCACCAGGTCTACAACATTACCTGGGAAGTGACCAATGGG 5936

      66 AspThrGlyValThrValAsnSerThrArgGlyValAlaProArgGlyTh 82
         |||         ||||||   :::      |||       |||       ||
    5937 GATCGG...GAGACAGTATGGGCAATATCAGGCAACCACCCTCTGTGGAC 5983

      82 rTrpTrpProGluLeu............................... 87
         ||||||||||   |||
    5984 TTGGTGGCCAGTCCTCACCCCAGATTTGTGTATGTTAGCTCTCAGTGGGC 6033

      88 .....HisPhe.................................... 89
              ||:::
    6034 CGCCCCACTGGGGGCTAGAGTATCAGGCCCCCTATTCCTCGCCCCCGGGG 6083

      90 .....................................CysLeuArgLe 93
                                             |||   |||
    6084 CCCCCTTGTTGCTCAGGGAGCAGCGGGAACGTTGCAGGCTGTGCCAGAGA 6133

      93 uIleAsnProAlaValLysSerThrProPro.................A 104
            |||      :::   |||       |||                 |
    6134 CTGCAACGAGCCCTTGACCTCCCTCACCCCTCGGTGCAACACTGCCTGGA 6183

     104 snLeuValArg.......................SerTyrGlyPheTyr 112
         ||   ::::::                      |||   |||||||||
    6184 ACAGACTTAAGCTGGACCAGGTAACTCATAAATCAAGTGAGGGATTTTAT 6233

     113 CysCysProGlyThrGluLys.........GluLysTyrCysGlyGlySe 126
            |||||||||:::::::            |||   |||||||||
    6234 GTCTGCCCCGGGTCACATCGCCCCCGGGAAGCCAAGTCCTGTGGGGGTCC 6283

     126 rGlyGluSerPheCysArgArgTrpSerCysValThrSerAsnAspGlyA 143
            :::   :::|||      |||:::|||   |||::::::
    6284 AGACTCCTTCTACTGTGCCTCTTGGGGCTGCGAGACAACCGGTAGAGTAT 6333

     143 spTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159
           ||||||   |||   |||      |||   :::              |||
    6334 ACTGGAAG...CCCTCCTCTTCTTGGGACTACATCACA.........GTA 6371

     160 AsnSerGlyProGlyLysTyrLysMetMetLysLeuTyrLysAsp...Ly 175
         :::::::::      :::   :::   :::::::::   ||||||   ||
    6372 GACAACAATCTCACCTCTAACCAGGCTGTTCAGGTATGCAAAGACAATAA 6421

     175 sSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluL 192
         |   |||:::|||                |||   |||   |||||||:::
    6422 GTGGTGCAATCCC..............TTGGCTATCCGGTTTACAAACG 6456

     192 ysGlyLysGlnGluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGly 208
           |||||||||      :::   :::|||   :::|||      ||||||
    6457 CCGGGAAACAG......GTCACCTCATGGACAACTGGACACTATTGGGGT 6500

     209 IleValPheTyrLysTyrGlyGlyGlyAlaGlySerThrLeuThrIleAr 225
         :::   :::|||      |||          |||   |||:::   |||||
    6501 CTACGTCTTTATGTCTCTGGACAGGACCCAGGGCTTACTTTCGGGATCCG 6550

     225 gLeuArgIleGluThrGlyThrGluProProValAlaMetGlyProAspL 242
         ||||.     :::   :::      |||   :::   :::|||||||:::
    6551 ACTCAGTTATCAA...AATCTAGGACCTCGGATCCCAATAGGACCAAACC 6597

     242 ysValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeu 258
           |||||||||:::|||            |||:::   |||
    6598 CCGTCCTGGCAGACCAA.............CTTTCGTTCCCGCTACCTAAT 6635

     259 ProValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSe 275
         |||:::|||                :::|||      :::   |||   ||
    6636 CCCCTACCC...............AAACCTGCCAAGTCTCCCCCCGCCTC 6670

     275 rAsnSerThrThrGlyLeuIle...ProThrAsnThrProArgAsnSerP 291
         |:::||||||      ||||||   |||:::   ||||||   :::   |
    6671 TAGTTCGACTCCCACATTGATTTCCCCGTCCCCCACTCCCACTCAGCCCC 6720

     291 roGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGly 307
         ||      |||:::   |||||||:::||||||:::::|||:::||||||
    6721 CG......CCAGCAGGAACGGGAGACAGATTACTAAATCTAGTACAGGGA 6764

     308 AlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTr 324
         |||:::||||||:::|||   |||:::||||||   |||:::::::|||||
    6765 GCTTACCAGGCACTCAACCTTACCAACCCTGATAAAACTCAAGAGTGCTG 6814

     324 pLeuCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluA 341
         ||||||||||   |||||||||||||||||||||||||:::|||
    6815 GTTATGCCTAGTGTCTGGACCCCCCTATTACGAGGGGGTTGCCGTCCTAG 6864

     341 rgLysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySer 357
            ::::::          :::          :::|||:::   :::|||
    6865 GTACTTATTCCAACCATACCTCTGCCCCAGCTAACTGCTCCGTGGCCTCC 6914

     358 ArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGl 374
         ::::::||||||||||||:::||||||:::|||:::|||   ||||||||
    6915 CAACACAAGCTGACCCTGTCCGAAGTGACTGGACGGGGACTCTGCATAGG 6964

     374 yLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrG 391
         |   :::|||   :::||||||   ||||||   :::|||:::      :
    6965 AACAGTCCCAAAAACTCACCAGGCCCTGTGCAACACTACCCTT......A 7008

     391 luGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrp 407
         ::   ::::::   :::   |||||||||          :::      |||
    7009 AGGCAGGCAAAGGGTCTTACTATCTAGTTGCCCCCACAGGAACTATGTGG 7058

     408 AlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGl 424
         ||||||||||||||||||||||||||:::|||::::::|||:::|||::
    7059 GCATGTAACACTGGACTCACTCCATGCCTATCTGCCACCGTGCTTAATCG 7108

     424 nSerLysAspLeuCysValMetValGlnIleValProArgValTyrTyrH 441
         ::::   |||   ||||||:::|||::::::   |||||||||   ||||
    7109 CACCACTGACTATTGCGTTCTCGTGGAATTATGGCCCAGGGTCACCTACC 7158

     441 isProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgPro 457
         |||||   :::   |||   ::::::::::   :::   :::|||
    7159 ATCCTCCCAGTTACGTCTATAGCCAGTTTGAA...AAATCCCATAGACAT 7205

     458 LysArgGluProValSerLeuThrLeuAlaValMetLeuGly........ 471
         ||||||||||||||||||||||||||||||:::::||||||
    7206 AAAAGAGAACCAGTGTCCTTAACCTTGGCCTTATTATTAGGTGGGCTAAC 7255

     472 ....LeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleT 487
                 |||   |||:::|||||||||||||||||:::||||||:::
    7256 TATGGGTGGCATCGCCGCGGGAGTAGGGACAGGAACTACCGCCCTGGTC. 7304

     487 hrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMet 503
           :::   ||||||:::::            :::|||||||||||||:::
    7305 ..GCCACCCAGCAGTTTCAG............CAGCTCCATGCTGCCGTA 7340

     504 ThrGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSe 520
            :::||||||:::   :::::::::|||::::::||||||||||:::||
    7341 CAAGATGATCTCAAAGAAGTCGAAAAGTCAATTACTAACCTAGAAAAGTC 7390

     520 rLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspL 537
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    7391 TCTTACTTCGTTGTCTGAGGTTGTACTGCAGAATCGACGAGGCCTAGACC 7440

     537 euLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCys 553
```

```
          ||||||||||:::|||||||||||||||||||||||||||||||||||
     7441 TGTTGTTCCTAAAAGAGGGAGGACTGTGTGCTGCCCTAAAAGAAGAATGT 7490

      554 CysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLe 570
          |||||||||:::||||||:::|||   :::||||||||||||   |||||
     7491 TGTTTCTATGCTGACCATACAGGCCTAGTAAGAGATAGTATGGCCAAATT 7540

      570 uArgLysLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyT 587
          ||||::::::|||::::::|||::::::      |||:::::::||||||
     7541 AAGAGAGAGACTCTCTCAGAGACAAAAACTATTTGAGTCGAGCCAAGGAT 7590

      587 rpPheGluGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSer 603
          ||||||||||||||||||||||||||||||||:::|||||||||:::|||
     7591 GGTTCGAAGGATGGTTTAACAGATCCCCCTGGTTTACCACGTTGATATCC 7640

      604 AlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThrValGlyPr 620
          ::::::   |||||||||:::::|||||||||:::|||      |||||
     7641 ACCATCATGGGGCCTCTCATTATACTCCTACTAATTCTGCTTTTTGGACC 7690

      620 oCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaV 637
          ||||::::::||||||:::|||   ||||||:::::|||:::|||:::|
     7691 CTGCATTCTTAATCGATTAGTTCAATTTGTTAAAGACAGGATCTCAGTAG 7740

      637 alGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 649
          |||||   :::||||||   |||||||||:::   |||
     7741 TCCAGGCTTTAGTCCTGACTCAACAATACCACCAGCTA 7778

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-8

seq_documentation_block:
; Sequence 8, Application PC/TUS9308041
;  GENERAL INFORMATION:
;    APPLICANT:  The Public Health Research Institute of the City of
;    APPLICANT:  New York, Inc.
;    TITLE OF INVENTION:  FUSION GLYCOPROTEINS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Davis Hoxie Faithfull and Hapgood
;      STREET:  45 Rockefeller Pl.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08041
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/938,100
;      FILING DATE:  28-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/_____
;      FILING DATE:  20-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jacobs, Seth H.
;      REGISTRATION NUMBER:  32,140
;      REFERENCE/DOCKET NUMBER:  11698A50
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-757-2200
;      TELEFAX:  212-586-1461
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-08041-8


alignment_scores:
            Quality: 1329.00              Length:    696
              Ratio:  2.973                Gaps:     23
 Percent Similarity:  64.224   Percent Identity: 42.816

alignment_block:
 US-09-171-553B-10 x PCT-US93-08041-8   ..

 Align seg 1/1  to: PCT-US93-08041-8  from: 1  to: 8323

       16 GluProLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPh 32
          :::|||:::   |||   |||||||||   :::                ||
     5817 GACCCGCGGGACCTCCTAATCCCCTTAATTCTC.............TT 5851

       32 eLeuThrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspS 49
          ||||.:::|||                  :::::::::          :
     5852 CCTGTCTCTC.............AAAGGGGCCAGATCCGCAGCACCCG 5886

       49 erSerAsnProHisArgProLeuSerProThrTrpLeuIleIleAspPro 65
          ::|||:::||||||:::      :::   ||||||   :::   :::
     5887 GCTCCAGCCCTCACCAGGTCTACAACATTACCTGGGAAGTGACCAATGGG 5936

       66 AspThrGlyValThrValAsnSerThrArgGlyValAlaProArgGlyTh 82
          |||         ||||||   :::   |||      |||      ||
     5937 GATCGG...GAGACAGTATGGGCAATATCAGGCAACCACCCTCTGTGGAC 5983

       82 rTrpTrpProGluLeu..................................... 87
          ||||||||||   |||
     5984 TTGGTGGCCAGTCCTCACCCCAGATTTGTGTATGTTAGCTCTCAGTGGGC 6033

       88 .....HisPhe.......................................... 89
               |||:::
     6034 CGCCCCACTGGGGGCTAGAGTATCAGGCCCCCTATTCCTCGCCCCCGGGG 6083

       90 ..........................................CysLeuArgLe 93
                                                   |||   |||
     6084 CCCCCTTGTTGCTCAGGGAGCAGCGGGAACGTTGCAGGCTGTGCCAGAGA 6133

       93 uIleAsnProAlaValLysSerThrProPro.................A 104
             |||      :::   |||      |||                  |
     6134 CTGCAACGAGCCCTTGACCTCCCTCACCCCTCGGTGCAACACTGCCTGGA 6183

      104 snLeuValArg.......................SerTyrGlyPheTyr 112
          ||   ::::::                     |||   |||||||||
     6184 ACAGACTTAAGCTGGACCAGGTAACTCATAAATCAAGTGAGGGATTTTAT 6233

      113 CysCysProGlyThrGluLys.........GluLysTyrCysGlyGlySe 126
          |||||||||:::::::::            |||   ||||||||
     6234 GTCTGCCCCGGGTCACATCGCCCCCGGGAAGCCAAGTCCTGTGGGGGTCC 6283

      126 rGlyGluSerPheCysArgArgTrpSerCysValThrSerAsnAspGlyA 143
             :::   :::|||      |||:::|||   |||::::::
     6284 AGACTCCTTCTACTGTGCCTCTTGGGGCTGCGAGACAACCGGTAGAGTAT 6333

      143 spTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159
          ||||||   |||   |||      |||   :::            |||
     6334 ACTGGAAG...CCCTCCTCTTCTTGGGACTACATCACA.........GTA 6371

      160 AsnSerGlyProGlyLysTyrLysMetMetLysLeuTyrLysAsp...Ly 175
          :::::::::      :::   :::   :::::::::   ||||||   ||
     6372 GACAACAATCTCACCTCTAACCAGGCTGTTCAGGTATGCAAAGACAATAA 6421

      175 sSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluL 192
          |   |||:::|||               |||   |||   ||||||:::
     6422 GTGGTGCAATCCC.............TTGGCTATCCGGTTTACAAACG 6456

      192 ysGlyLysGlnGluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGly 208
          |||||||||      :::   :::|||   :::|||      ||||||
     6457 CCGGGAAACAG......GTCACCTCATGGACAACTGGACACTATTGGGGT 6500
```

```
 209 IleValPheTyrLysTyrGlyGlyGlyAlaGlySerThrLeuThrIleAr 225
6501 CTACGTCTTTATGTCTCTGGACAGGACCCAGGGCTTACTTTCGGGATCCG 6550

 225 gLeuArgIleGluThrGlyThrGluProProValAlaMetGlyProAspL 242
6551 ACTCAGTTATCAA...AATCTAGGACCTCGGATCCCAATAGGACCAAACC 6597

 242 ysValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeu 258
6598 CCGTCCTGGCAGACCAA.....CTTTCGTTCCCGCTACCTAAT 6635

 259 ProValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSe 275
6636 CCCCTACCC.....AAACCTGCCAAGTCTCCCCCCGCCTC 6670

 275 rAsnSerThrThrGlyLeuIle...ProThrAsnThrProArgAsnSerP 291
6671 TAGTTCGACTCCCACATTGATTTCCCCGTCCCCCACTCCCACTCAGCCCC 6720

 291 roGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGly 307
6721 CG......CCAGCAGGAACGGGAGACAGATTACTAAATCTAGTACAGGGA 6764

 308 AlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTr 324
6765 GCTTACCAGGCACTCAACCTTACCAACCCTGATAAAACTCAAGAGTGCTG 6814

 324 pLeuCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluA 341
6815 GTTATGCCTAGTGTCTGGACCCCCCTATTACGAGGGGGTTGCCGTCCTAG 6864

 341 rgLysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySer 357
6865 GTACTTATTCCAACCATACCTCTGCCCCAGCTAACTGCTCCGTGGCCTCC 6914

 358 ArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGl 374
6915 CAACACAAGCTGACCCTGTCCGAAGTGACTGGACGGGGACTCTGCATAGG 6964

 374 yLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrG 391
6965 AACAGTCCCAAAAACTCACCAGGCCCTGTGCAACACTACCCTT......A 7008

 391 luGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrp 407
7009 AGGCAGGCAAAGGGTCTTACTATCTAGTTGCCCCCACAGGAACTATGTGG 7058

 408 AlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGl 424
7059 GCATGTAACACTGGACTCACTCCATGCCTATCTGCCACCGTGCTTAATCG 7108

 424 nSerLysAspLeuCysValMetValGlnIleValProArgValTyrTyrH 441
7109 CACCACTGACTATTGCGTTCTCGTGGAATTATGGCCCAGGGTCACCTACC 7158

 441 isProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgPro 457
7159 ATCCTCCCAGTTACGTCTATAGCCAGTTTGAA...AAATCCCATAGACAT 7205

 458 LysArgGluProValSerLeuThrLeuAlaValMetLeuGly........ 471
7206 AAAAGAGAACCAGTGTCCTTAACCTTGGCCTTATTATTAGGTGGGCTAAC 7255

 472 ....LeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleT 487
7256 TATGGGTGGCATCGCCGCGGGAGTAGGGACAGGAACTACCGCCCTGGTC. 7304

 487 hrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMet 503
7305 ..GCCACCCAGCAGTTTCAG.....CAGCTCCATGCTGCCGTA 7340

 504 ThrGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSe 520
7341 CAAGATGATCTCAAAGAAGTCGAAAAGTCAATTACTAACCTAGAAAAGTC 7390

 520 rLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspL 537
7391 TCTTACTTCGTTGTCTGAGGTTGTACTGCAGAATCGACGAGGCCTAGACC 7440

 537 euLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCys 553
7441 TGTTGTTCCTAAAAGAGGGAGGACTGTGTGCTGCCCTAAAAGAAGAATGT 7490

 554 CysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLe 570
7491 TGTTTCTATGCTGACCATACAGGCCTAGTAAGAGATAGTATGGCCAAATT 7540

 570 uArgLysLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyT 587
7541 AAGAGAGAGACTCTCTCAGAGACAAAAACTATTTGAGTCGAGCCAAGGAT 7590

 587 rpPheGluGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSer 603
7591 GGTTCGAAGGATGGTTTAACAGATCCCCCTGGTTTACCACGTTGATATCC 7640

 604 AlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThrValGlyPr 620
7641 ACCATCATGGGGCCTCTCATTATACTCCTACTAATTCTGCTTTTTGGACC 7690

 620 oCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaV 637
7691 CTGCATTCTTAATCGATTAGTTCAATTTGTTAAAGACAGGATCTCAGTAG 7740

 637 alGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 649
7741 TCCAGGCTTTAGTCCTGACTCAACAATACCACCAGCTA 7778
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-6

```
seq_documentation_block:
; Sequence 6, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
;  APPLICANT: Collins, Mary KL
;  APPLICANT: Weiss, Robin A
;  APPLICANT: Takeuchi, Yasuhiro
;  APPLICANT: Cosset, Francois-Loic
;  TITLE OF INVENTION: Expression systems
;  FILE REFERENCE: 09/011,745
;  CURRENT APPLICATION NUMBER: US/09/011,745
;  CURRENT FILING DATE: 1998-06-22
;  EARLIER APPLICATION NUMBER: PCT/GB96/02061
;  EARLIER FILING DATE: 1996-08-23
;  EARLIER APPLICATION NUMBER: GB9517263.1
;  EARLIER FILING DATE: 1995-08-23
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 6061
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Portion of
;   OTHER INFORMATION: construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3807)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3808)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (3809)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3810)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3995)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3996)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3997)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3998)
;   OTHER INFORMATION: n is any nucleotide
US-09-011-745-6


alignment_scores:
            Quality: 1284.00              Length:    734
              Ratio:  2.828                Gaps:     26
 Percent Similarity:  61.853   Percent Identity: 41.008

alignment_block:
 US-09-171-553B-10 x US-09-011-745-6   ..

 Align seg 1/1  to: US-09-011-745-6  from: 1  to: 6061

       2 HisProThrLeuSerTrpArgHisLeuProThrArgGlyGlyGluProLy 18
         |||   :::|||:::   |||:::   |||   :::|||   :::|||::
    1286 CATCGCAGCTTGGATACACGCCGC...CCACGTGAAGGCTGCCGACCCCG 1332

      18 sArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrL 35
         :                                  :::|||   :::
    1333 G...............................GGGTGGACCATCCTCT 1349

      35 euThrIleThrProGlnAlaSer...SerLysArgLeuIleAspSerSer 50
           |||         :::::::::   ||||||   |||   ::::::
    1350 AGACTGACATGGCGCGTTCAACGCTCTCAAAACCCCTTAAAAATAAGGTT 1399

      51 AsnProHisArgProLeuSerPro.......................... 58
         ||||||:::   ||||||   |||
    1400 AACCCGCGAGGCCCCCTAATCCCCTTAATTCTTCTGATGCTCAGAGGGGT 1449

      59 ..............................................ThrT 60
                                                      ||||
    1450 CAGTACTGCTTCGCCCGGCTCCAGTCCTCATCAAGTCTATAATATCACCT 1499

      60 rpLeuIleIleAspProAspThrGlyValThrValAsnSerThrArgGly 76
         ||   :::   :::   |||         ||||||   :::|||   |||
    1500 GGGAGGTAACCAATGGAGATCGG...GAGACGGTATGGGCAACTTCTGGC 1546

      77 ValAlaProArgGlyThrTrpTrpProGluLeu.................. 87
           |||      ||||||||||||:::|||
    1547 AACCACCCTCTGTGGACCTGGTGGCCTGACCTTACCCCAGATTTATGTAT 1596

      88 .....................HisPheCysLeuArgLeuIleAsnProA 97
                              ::::::   |||:::      :::|||
    1597 GTTAGCCCACCATGGACCATCTTATTGGGGGCTAGAATATCAATCCCCTT 1646

      97 laValLysSerThrProPro............................. 103
               |||:::||||||
    1647 TT......TCTTCTCCCCCGGGGCCCCCTTGTTGCTCAGGGGGCAGCAGC 1690

     103 ............................................... 103
```

```
1691 CCAGGCTGTTCCAGAGACTGCGAAGAACCTTTAACCTCCCTCACCCCTCG 1740

 104 ...............AsnLeuValArg..................... 107
                       |||   ::::::
1741 GTGCAACACTGCCTGGAACAGACTCAAGCTAGACCAGACAACTCATAAAT 1790

 108 ..SerTyrGlyPheTyrCysCysProGlyThrGluLys.........Glu 120
       :::   |||||||||   |||||||||   ::::::          :::
1791 CAAATGAGGGATTTTATGTTTGCCCCGGGCCCCACCGCCCCCGAGAATCC 1840

 121 LysTyrCysGlyGlySerGlyGluSerPheCysArgArgTrpSerCysVa 137
       |||   |||||||||       :::   :::|||       |||:::|||
1841 AAGTCATGTGGGGGTCCAGACTCCTTCTACTGTGCCTATTGGGGCTGTGA 1890

 137 lThrSerAsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgV 154
       |||::::::   :::   ||||||   |||   |||       |||
1891 GACAACCGGTAGAGCTTACTGGAAG...CCCTCCTCATCATGGGAT.... 1933

 154 alLysPheSerPheValAsnSerGlyProGlyLysTyrLysMetMetLys 170
           |||      ||||||::::::      :::   :::   ::::::
1934 .....TTCATCACAGTAAACAACAATCTCACCTCTGACCAGGCTGTCCAG 1978

 171 LeuTyrLysAsp...LysSerCysSerProSerAspLeuAspTyrLeuLy 186
       :::   ||||||   |||   |||:::|||                |||
1979 GTATGCAAAGATAATAAGTGGTGCAACCCC..............TTAGT 2013

 186 sIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrpIleA 203
       |||   ||||||:::   |||::::::      :::   :::|||   :
2014 TATTCGGTTTACAGACGCCGGGAGACGG......GTTACTTCCTGGACCA 2057

 203 snGlyMetSerTrpGlyIleValPheTyrLysTyrGlyGlyGlyAlaGly 219
       ::|||      ||||||:::   :::|||      |||         |||
2058 CAGGACATTACTGGGGCTTACGTTTGTATGTCTCCGGACAAGATCCAGGG 2107

 220 SerThrLeuThrIleArgLeuArgIleGluThrGlyThrGluProProVa 236
       |||:::   ||||||||||||   :::   :::      |||   ||
2108 CTTACATTTGGGATCCGACTCAGATACCAA...AATCTAGGACCCCGCGT 2154

 236 lAlaMetGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeuG 253
       |   :::|||||:::   |||||||||:::|||   |||      |||:
2155 CCCAATAGGGCCAAACCCCGTTCTGGCAGACCAACAGCCA......CTCT 2198

 253 luProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAsp 269
       ::   |||         ||||||   :::                |||:::
2199 CCAAGCCCAAA......CCTGTTAAGTCG..............CCTTCA 2227

 270 IleThrGlnProProSerAsnSerThrThrGlyLeuIleProThrAsnTh 286
       :::|||:::|||||||||:::      |||   |||   |||||||:::
2228 GTCACCAAACCACCCAGTGGG......ACTCCTCTCTCCCCTACCCAACT 2271

 286 rProArgAsnSerProGlyValProValLysThrGlyGlnArgLeuPheS 303
       |||                  |||:::   |||   :::|||||:::::
2272 TCCA..................CCGGCGGGAACGGAAAATAGGCTGCTAA 2303

 303 erLeuIleGlnGlyAlaPheGlnAlaIleAsnSerThrAspProAspAla 319
       ::|||:::::|||||||:::|||||||:::|||   |||:::||||||
2304 ACTTAGTAGACGGAGCCTACCAAGCCCTCAACCTCACCAGTCCTGACAAA 2353

 320 ThrSerSerCysTrpLeuCysLeuSerSerGlyProProTyrTyrGluGl 336
       |||::::::|||||||||||||||   :::|||||||||||||||||||
2354 ACCCAAGAGTGCTGGTTGTGTCTAGTAGCGGGACCCCCCTACTACGAAGG 2403

 336 yMetAlaLysGluArgLysPheAsnValThrLysGluHisArgAsnGlnC 353
       |:::|||      ::::::         :::          :::|
2404 GGTTGCCGTCCTGGGTACCTACTCCAACCATACCTCTGCTCCAGCCAACT 2453

 353 ysThrTrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLys 369
       ||:::   :::|||:::::|||||||||||:::|||||||:::|||:::
2454 GCTCCGTGGCCTCCCAACACAAGTTGACCCTGTCCGAAGTGACCGGACAG 2503

 370 GlyThrCysIleGlyLysAlaProProSerHisGlnHisLeuCysTyrSe 386
       |||   |||||||||   :::|||   :::||||||   ||||||   ::
2504 GGACTCTGCATAGGAGCAGTTCCCAAAACACATCAGGCCCTATGTAATAC 2553
```

386 rThrValValTyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyT 403
2554 CACCCAGACAAGCAGTCGAGGGTCC.....TATTATCTAGTTGCCCCTA 2597
403 yrAsnArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThr 419
2598 CAGGTACCATGTGGGCTTGTAGTACCGGGCTTACTCCATGCATCTCCACC 2647
420 SerValPheAsnGlnSerLysAspLeuCysValMetValGlnIleValPr 436
2648 ACCATACTGAACCTTACCACTGATTATTGTGTTCTTGTCGAACTCTGGCC 2697
436 oArgValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyrA 453
2698 AAGAGTCACCTATCATTCCCCCAGCTATGTTTACGGCCTGTTTGAG...A 2744
453 rgTyrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMet 469
2745 GATCCAACCGACACAAAAGAGAACCGGTGTCGTTAACCCTGGCCCTATTA 2794
470 LeuGly.....LeuGlyThrAlaValGlyValGlyThrGlyTh 482
2795 TTGGGTGGACTAACCATGGGGGGAATTGCCGCTGGAATAGGAACAGGGAC 2844
482 rAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyGluL 499
2845 TACTGCTCTAATG...GCCACTCAGCAATTCCAG.....CAGC 2879
499 euHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSer 515
2880 TCCAAGCCGCAGTACAGGATGATCTCAGGGAGGTTGAAAAATCAATCTCT 2929
516 AsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnAr 532
2930 AACCTAGAAAAGTCTCTCACTTCCCTGTCTGAAGTTGTCCTACAGAATCG 2979
532 gArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaL 549
2980 AAGGGGCCTAGACTTGTTATTTCTAAAAGAAGGAGGGCTGTGTGCTGCTC 3029
549 euLysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAsp 565
3030 TAAAAGAAGAATGTTGCTTCTATGCGGACCACACAGGACTAGTGAGAGAC 3079
566 SerMetAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArgGl 582
3080 AGCATGGCCAAATTGAGAGAGAGGCTTAATCAGAGACAGAAACTGTTTGA 3129
582 uAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMetT 599
3130 GTCAACTCAAGGATGGTTTGAGGGACTGTTTAACAGATCCCCTTGGTTTA 3179
599 hrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeu 615
3180 CCACCTTGATATCTACCATTATGGGACCCCTCATTGTACTCCTAATGATT 3229
616 LeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGl 632
3230 TTGCTCTTCGGACCCTGCATTCTTAATCGATTAGTTCAATTTGTTAAAGA 3279
632 uArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyL 649
3280 CAGGATCTCAGTAGTCCAGGCTTTAGTCCTGACTCAACAATACCACCAGC 3329
649 eu 649
3330 TA 3331

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 03:01:22 ; Search time 212.72 Seconds
(without alignments)
21.294 Million cell updates/sec

Title: US-09-171-553B-13
Perfect score: 20
Sequence: 1 gatggctctcctgccctttg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	20	100.0	7333	4	US-08-766-528-2	Sequence 2, Appli
	2	18.4	92.0	3482	3	US-09-111-085-3	Sequence 3, Appli
	3	18.4	92.0	8060	4	US-08-766-528-1	Sequence 1, Appli
	4	18.4	92.0	8132	4	US-08-766-528-3	Sequence 3, Appli
c	5	16.8	84.0	72928	3	US-09-009-913-1	Sequence 1, Appli
c	6	16	80.0	23	4	US-08-766-528-14	Sequence 14, Appl
	7	16	80.0	23	4	US-08-766-528-15	Sequence 15, Appl
c	8	16	80.0	2035	5	PCT-US93-12464-1	Sequence 1, Appli
	9	16	80.0	2400	3	US-08-800-291B-7	Sequence 7, Appli
	10	15.4	77.0	4060	1	US-08-308-949A-1	Sequence 1, Appli
c	11	15.4	77.0	31571	1	US-08-323-443B-1	Sequence 1, Appli
c	12	15.4	77.0	53526	3	US-08-658-136-2	Sequence 2, Appli
c	13	15.4	77.0	53577	3	US-08-658-136-1	Sequence 1, Appli
	14	15.2	76.0	3103	3	US-08-826-246-3	Sequence 3, Appli
	15	15.2	76.0	3103	3	US-08-944-495-3	Sequence 3, Appli
	16	15.2	76.0	3103	3	US-09-126-640-2	Sequence 2, Appli
	17	15.2	76.0	3103	4	US-08-925-588-3	Sequence 3, Appli
	18	15.2	76.0	3111	3	US-09-487-444-3	Sequence 3, Appli
	19	15.2	76.0	8655	3	US-09-075-272-1	Sequence 1, Appli
	20	15.2	76.0	15101	2	US-08-799-464A-14	Sequence 14, Appl
	21	15.2	76.0	15101	5	PCT-US95-09927-14	Sequence 14, Appl
	22	15.2	76.0	15108	1	US-08-157-005-1	Sequence 1, Appli
	23	15.2	76.0	15108	4	US-08-747-863-1	Sequence 1, Appli
c	24	15	75.0	68	5	PCT-US95-17026-8	Sequence 8, Appli
	25	15	75.0	70	5	PCT-US95-17026-9	Sequence 9, Appli
c	26	15	75.0	3876	5	PCT-US95-17026-1	Sequence 1, Appli
c	27	14.8	74.0	489	1	US-08-081-072-9	Sequence 9, Appli
c	28	14.8	74.0	489	1	US-08-449-093A-9	Sequence 9, Appli
c	29	14.8	74.0	569	2	US-08-483-695-44	Sequence 44, Appl
c	30	14.8	74.0	569	2	US-07-965-285-44	Sequence 44, Appl
c	31	14.8	74.0	569	2	US-08-487-231-44	Sequence 44, Appl
c	32	14.8	74.0	569	4	US-09-201-912-44	Sequence 44, Appl
	33	14.8	74.0	7886	2	US-08-751-189-2	Sequence 2, Appli
	34	14.8	74.0	7886	2	US-09-060-836-2	Sequence 2, Appli
	35	14.8	74.0	7886	4	US-09-184-445-2	Sequence 2, Appli
c	36	14.4	72.0	1539	2	US-08-828-596-1	Sequence 1, Appli
c	37	14.4	72.0	1666	4	US-09-360-197-5	Sequence 5, Appli
c	38	14.4	72.0	2567	3	US-08-993-260-4	Sequence 4, Appli
	39	14.2	71.0	1160	2	US-08-955-138-5	Sequence 5, Appli
c	40	14.2	71.0	1470	6	5187079-3	Patent No. 5187079
c	41	14.2	71.0	3083	1	US-08-346-849-1	Sequence 1, Appli
c	42	14.2	71.0	3083	2	US-08-293-284A-1	Sequence 1, Appli
	43	14.2	71.0	4020	4	US-09-050-159-130	Sequence 130, App
	44	14.2	71.0	4024	4	US-09-162-484-18	Sequence 18, Appl
c	45	14.2	71.0	4928	1	US-08-399-561-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-766-528-2
; Sequence 2, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-2


  Query Match             100.0%;  Score 20;  DB 4;  Length 7333;
  Best Local Similarity   100.0%;  Pred. No. 0.85;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gatggctctcctgccctttg 20
           ||||||||||||||||||||
```

```
Db   4442 GATGGCTCTCCTGCCCTTTG 4461


RESULT   2
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
;  APPLICANT: Stoye, Jonathan P
;  APPLICANT: Weiss, Robin A
;  TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
;  TITLE OF INVENTION: specific sequences
;  FILE REFERENCE: 4238/75168
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 3482
;   TYPE: DNA
;   ORGANISM: Porcine retrovirus
US-09-111-085-3


  Query Match             92.0%;  Score 18.4;  DB 3;  Length 3482;
  Best Local Similarity   95.0%;  Pred. No. 4.3;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gatggctctcctgccctttg 20
           ||| ||||||||||||||||
Db     528 gatagctctcctgccctttg 547


RESULT   3
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1


  Query Match             92.0%;  Score 18.4;  DB 4;  Length 8060;
  Best Local Similarity   95.0%;  Pred. No. 4.7;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gatggctctcctgccctttg 20
           ||| ||||||||||||||||
Db    7763 GATAGCTCTCCTGCCCTTTG 7782


RESULT   4
US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8132 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-3


  Query Match             92.0%;  Score 18.4;  DB 4;  Length 8132;
  Best Local Similarity   95.0%;  Pred. No. 4.7;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gatggctctcctgccctttg 20
           ||| ||||||||||||||||
Db    5237 GATAGCTCTCCTGCCCTTTG 5256


RESULT   5
US-09-009-913-1/c
; Sequence 1, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  AxyS Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  Asthma Related Genes
;    NUMBER OF SEQUENCES:  339
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  SEQ-4P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3231
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  72928 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-09-009-913-1


  Query Match             84.0%;  Score 16.8;  DB 3;  Length 72928;
  Best Local Similarity   90.0%;  Pred. No. 32;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccctttg 20
          |||||||| | |||||||||
Db  46404 GATGGCTCCCATGCCCTTTG 46385


RESULT   6
US-08-766-528-14/c
; Sequence 14, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-14


  Query Match             80.0%;  Score 16;  DB 4;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 gctctcctgccctttg 20
            ||||||||||||||||
Db      23 GCTCTCCTGCCCTTTG 8


RESULT   7
US-08-766-528-15
; Sequence 15, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-15
```

```
  Query Match           80.0%;  Score 16;  DB 4;  Length 23;
  Best Local Similarity  100.0%;  Pred. No. 34;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 gctctcctgccctttg 20
          ||||||||||||||||
Db      1 GCTCTCCTGCCCTTTG 16


RESULT   8
PCT-US93-12464-1/c
; Sequence 1, Application PC/TUS9312464
;  GENERAL INFORMATION:
;    APPLICANT:  The Regents of the University of California, et al.
;    TITLE OF INVENTION:  METHOD FOR INHIBITION OF CELL
;    TITLE OF INVENTION:  ADHESION TO RECEPTORS CONTAINING SELECTINS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPENSLEY HORN JUBAS & LUBITZ
;      STREET:  1880 CENTURY PARK EAST, FIFTH FLOOR
;      CITY:  LOS ANGELES
;      STATE:  CALIFORNIA
;      COUNTRY:  US
;      ZIP:  90067
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/12464
;      FILING DATE:  21-DEC-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HOWELLS, STACY L.
;      REGISTRATION NUMBER:  34,842
;      REFERENCE/DOCKET NUMBER:  FD2929
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 455-5100
;      TELEFAX:  (619) 455-5110
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  ALPHA-L-FUCOSIDASE
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..1401
PCT-US93-12464-1


  Query Match           80.0%;  Score 16;  DB 5;  Length 2035;
  Best Local Similarity  100.0%;  Pred. No. 53;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 gctctcctgccctttg 20
          ||||||||||||||||
Db   1851 GCTCTCCTGCCCTTTG 1836


RESULT   9
US-08-800-291B-7
; Sequence 7, Application US/08800291B
; Patent No. 6153740
;  GENERAL INFORMATION:
;    APPLICANT:  J.D. Young & C.E. Cass
;    TITLE OF INVENTION:  cDNA ENCODING NUCLEOSIDE TRANSPORTER
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/800,291B
;      FILING DATE:  13-FEB-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/499,314
;      FILING DATE:  7-JULY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07254/044WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2400 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2401
US-08-800-291B-7


  Query Match           80.0%;  Score 16;  DB 3;  Length 2400;
  Best Local Similarity  100.0%;  Pred. No. 53;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 gctctcctgccctttg 20
          ||||||||||||||||
Db   2256 GCTCTCCTGCCCTTTG 2271


RESULT  10
US-08-308-949A-1
; Sequence 1, Application US/08308949A
; Patent No. 5580703
;  GENERAL INFORMATION:
;    APPLICANT:  Kotin, Robert M.
;    APPLICANT:  Berns, Kenneth I.
;    APPLICANT:  Linden, Ralph M.
;    TITLE OF INVENTION:  Human Adeno-Associated Virus Integration
;    TITLE OF INVENTION:  Site DNA and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/308,949A
;      FILING DATE:  September 20, 1994
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/947,127
;      FILING DATE:  September 27, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Alice O.
;      REGISTRATION NUMBER:  33,542
;      REFERENCE/DOCKET NUMBER:  ACC92-10F
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-308-949A-1


  Query Match             77.0%;  Score 15.4;  DB 1;  Length 4060;
  Best Local Similarity   94.1%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 tggctctcctgcccttt 19
           |||||||||||||| ||
Db    2970 TGGCTCTCCTGCCCCTT 2986


RESULT  11
US-08-323-443B-1/c
; Sequence 1, Application US/08323443B
; Patent No. 5654170
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W.
;    APPLICANT:  LANDES, GREGORY M.
;    APPLICANT:  BURN, TIMOTHY C.
;    APPLICANT:  CONNORS, TIMOTHY D.
;    APPLICANT:  DACKOWSKI, WILLIAM R.
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/323,443B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, S. Peter
;      REGISTRATION NUMBER:  25,351
;      REFERENCE/DOCKET NUMBER:  0372/0A462
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 527-7700
;      TELEFAX:  (212) 753-6237
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31571 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  PKD1 GENOMIC
US-08-323-443B-1


  Query Match             77.0%;  Score 15.4;  DB 1;  Length 31571;
  Best Local Similarity   94.1%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gatggctctcctgccct 17
           | ||||||||||||||||
Db  13938 GCTGGCTCTCCTGCCCT 13922


RESULT  12
US-08-658-136-2/c
; Sequence 2, Application US/08658136
; Patent No. 6071717
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W
;    APPLICANT:  LANDES, GREGORY M
;    APPLICANT:  BURN, TIMOTHY C
;    APPLICANT:  CONNORS, TIMOTHY D
;    APPLICANT:  DACKOWSKI, WILLIAM
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  ONE MOUNTAIN ROAD
;      CITY:  FRAMINGHAM
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,136
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LASSEN, ELIZABETH
;      REGISTRATION NUMBER:  31,845
;      REFERENCE/DOCKET NUMBER:  GEN4-17.8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  508-872-8400
;      TELEFAX:  508-872-5415
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53526 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-658-136-2


  Query Match             77.0%;  Score 15.4;  DB 3;  Length 53526;
  Best Local Similarity   94.1%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gatggctctcctgccct 17
           | ||||||||||||||||
Db  14923 GCTGGCTCTCCTGCCCT 14907
```

```
RESULT  13
US-08-658-136-1/c
; Sequence 1, Application US/08658136
; Patent No. 6071717
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W
;    APPLICANT:  LANDES, GREGORY M
;    APPLICANT:  BURN, TIMOTHY C
;    APPLICANT:  CONNORS, TIMOTHY D
;    APPLICANT:  DACKOWSKI, WILLIAM
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  ONE MOUNTAIN ROAD
;      CITY:  FRAMINGHAM
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,136
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LASSEN, ELIZABETH
;      REGISTRATION NUMBER:  31,845
;      REFERENCE/DOCKET NUMBER:  GEN4-17.8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  508-872-8400
;      TELEFAX:  508-872-5415
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53577 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-658-136-1


  Query Match             77.0%;  Score 15.4;  DB 3;  Length 53577;
  Best Local Similarity   94.1%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccct 17
          | ||||||||||||||||
Db  14922 GCTGGCTCTCCTGCCCT 14906


RESULT  14
US-08-826-246-3
; Sequence 3, Application US/08826246
; Patent No. 6048709
;  GENERAL INFORMATION:
;    APPLICANT:  Falb, Dean
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION:  THE TREATMENT AND DIAGNOSIS OF
;    TITLE OF INVENTION:  CARDIOVASCULAR DISEASE
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/826,246
;      FILING DATE:  28-MAR-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/799,910
;      FILING DATE:  13-FEB-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/011,787
;      FILING DATE:  16-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-078-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)7909090
;      TELEFAX:  (212)8699741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3103 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  288...1565
;      OTHER INFORMATION:
US-08-826-246-3


  Query Match             76.0%;  Score 15.2;  DB 3;  Length 3103;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgcccttg 20
          | || ||||||||||| |||
Db   2031 GGTGTCTCTCCTGCCCCTTG 2050


RESULT  15
US-08-944-495-3
; Sequence 3, Application US/08944495
; Patent No. 6087477
;  GENERAL INFORMATION:
;    APPLICANT:  Falb, Dean
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION:  THE TREATMENT AND DIAGNOSIS OF
;    TITLE OF INVENTION:  CARDIOVASCULAR DISEASE
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/944,495
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  08/799,910
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  7853-067-999
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212)7909090
;       TELEFAX:  (212)8699741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3103 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  Coding Sequence
;       LOCATION:  288...1565
;       OTHER INFORMATION:
US-08-944-495-3


  Query Match             76.0%;  Score 15.2;  DB 3;  Length 3103;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gatggctctcctgcccttg 20
           | || |||||||||| |||
Db    2031 GGTGTCTCTCCTGCCCCTTG 2050


Search completed: February 24, 2002, 03:01:27
Job time: 21438 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 03:01:27 ; Search time 212.72 Seconds
(without alignments)
19.164 Million cell updates/sec

Title: US-09-171-553B-14
Perfect score: 18
Sequence: 1 ccacagtcgtacaccacg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	18	100.0	3482	3	US-09-111-085-3	Sequence 3, Appli
c	2	18	100.0	7333	4	US-08-766-528-2	Sequence 2, Appli
c	3	18	100.0	8060	4	US-08-766-528-1	Sequence 1, Appli
c	4	18	100.0	8132	4	US-08-766-528-3	Sequence 3, Appli
	5	17	94.4	20	4	US-08-766-528-61	Sequence 61, Appl
c	6	17	94.4	28	3	US-09-111-085-15	Sequence 15, Appl
	7	13.8	76.7	1760	3	US-09-150-133-8	Sequence 8, Appli
	8	13.8	76.7	1760	3	US-09-150-141-8	Sequence 8, Appli
	9	13.8	76.7	1760	4	US-09-374-493-8	Sequence 8, Appli
	10	13.8	76.7	1760	4	US-09-374-824-8	Sequence 8, Appli
	11	13.8	76.7	1760	4	US-09-374-492-8	Sequence 8, Appli
	12	13.8	76.7	2711	4	US-08-891-298-2	Sequence 2, Appli
c	13	13.8	76.7	3177	1	US-08-042-747A-4	Sequence 4, Appli
	14	13.8	76.7	5892	3	US-08-755-587-27	Sequence 27, Appl
c	15	13.8	76.7	6769	1	US-08-480-784-20	Sequence 20, Appl
c	16	13.8	76.7	6769	1	US-08-483-553-20	Sequence 20, Appl
c	17	13.8	76.7	6769	1	US-08-487-002-20	Sequence 20, Appl
c	18	13.8	76.7	6769	1	US-08-483-554B-20	Sequence 20, Appl
c	19	13.8	76.7	6769	1	US-08-488-011B-20	Sequence 20, Appl
c	20	13.8	76.7	6769	4	US-08-850-727-20	Sequence 20, Appl
c	21	13.8	76.7	6769	5	PCT-US95-10202-20	Sequence 20, Appl
c	22	13.8	76.7	6769	5	PCT-US95-10203-20	Sequence 20, Appl
c	23	13.8	76.7	6769	5	PCT-US95-10220-20	Sequence 20, Appl
	24	13.8	76.7	18073	4	US-09-078-294-12	Sequence 12, Appl
c	25	13.4	74.4	897	2	US-09-006-535-2	Sequence 2, Appli
c	26	13.4	74.4	897	2	US-09-006-535-7	Sequence 7, Appli
c	27	13.4	74.4	1320	1	US-08-419-414-1	Sequence 1, Appli
c	28	13.4	74.4	1525	2	US-09-006-535-1	Sequence 1, Appli
c	29	13.4	74.4	3058	4	US-08-818-112-106	Sequence 106, App
c	30	13.4	74.4	6485	3	US-09-081-320-2	Sequence 2, Appli
c	31	13.4	74.4	8743	3	US-09-081-320-1	Sequence 1, Appli
	32	13.4	74.4	36741	4	US-09-301-665-3	Sequence 3, Appli
c	33	13.4	74.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	34	13.2	73.3	187	2	US-08-485-657A-17	Sequence 17, Appl
c	35	13.2	73.3	187	5	PCT-US95-02303-25	Sequence 25, Appl
	36	13.2	73.3	188	2	US-08-485-657A-22	Sequence 22, Appl
	37	13.2	73.3	188	5	PCT-US95-02303-21	Sequence 21, Appl
c	38	13.2	73.3	1621	3	US-09-013-881-14	Sequence 14, Appl
c	39	13.2	73.3	15378	3	US-08-785-420-1	Sequence 1, Appli
c	40	13.2	73.3	40138	3	US-09-090-793-12	Sequence 12, Appl
	41	13.2	73.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	42	13	72.2	1342	3	US-08-961-083-181	Sequence 181, App
c	43	12.8	71.1	78	2	US-08-680-326-46	Sequence 46, Appl
	44	12.8	71.1	176	4	US-08-905-223-150	Sequence 150, App
c	45	12.8	71.1	560	3	US-08-545-809A-62	Sequence 62, Appl

ALIGNMENTS

```
RESULT   1
US-09-111-085-3/c
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
;  APPLICANT: Stoye, Jonathan P
;  APPLICANT: Weiss, Robin A
;  TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
;  TITLE OF INVENTION: specific sequences
;  FILE REFERENCE: 4238/75168
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 3482
;   TYPE: DNA
;   ORGANISM: Porcine retrovirus
US-09-111-085-3


  Query Match             100.0%;  Score 18;  DB 3;  Length 3482;
  Best Local Similarity   100.0%;  Pred. No. 0.72;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccacg 18
           ||||||||||||||||||
Db    3439 CCACAGTCGTACACCACG 3422


RESULT   2
US-08-766-528-2/c
; Sequence 2, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-2


  Query Match             100.0%;  Score 18;  DB 4;  Length 7333;
  Best Local Similarity   100.0%;  Pred. No. 0.74;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccacg 18
           ||||||||||||||||||
Db    7279 CCACAGTCGTACACCACG 7262


RESULT   3
US-08-766-528-1/c
; Sequence 1, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-1


  Query Match             100.0%;  Score 18;  DB 4;  Length 8060;
  Best Local Similarity   100.0%;  Pred. No. 0.75;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccacg 18
           ||||||||||||||||||
Db    2546 CCACAGTCGTACACCACG 2529


RESULT   4
US-08-766-528-3/c
; Sequence 3, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8132 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-3


  Query Match             100.0%;  Score 18;  DB 4;  Length 8132;
  Best Local Similarity   100.0%;  Pred. No. 0.75;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccacg 18
           ||||||||||||||||||
Db    8080 CCACAGTCGTACACCACG 8063


RESULT   5
US-08-766-528-61
; Sequence 61, Application US/08766528
; Patent No. 6190861
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Jay A. Fishman
;    TITLE OF INVENTION:  MOLECULAR SEQUENCE OF SWINE RETROVIRUS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,528
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/572,645
;      FILING DATE:  14-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  MGP-038CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  61:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-766-528-61


  Query Match             94.4%;  Score 17;  DB 4;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cacagtcgtacaccacg 18
          |||||||||||||||||
Db      1 CACAGTCGTACACCACG 17


RESULT   6
US-09-111-085-15/c
; Sequence 15, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
;  APPLICANT: Stoye, Jonathan P
;  APPLICANT: Weiss, Robin A
;  TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
;  TITLE OF INVENTION: specific sequences
;  FILE REFERENCE: 4238/75168
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 28
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:
;   OTHER INFORMATION: oligonucleotide
US-09-111-085-15


  Query Match             94.4%;  Score 17;  DB 3;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cacagtcgtacaccacg 18
          |||||||||||||||||
Db     28 CACAGTCGTACACCACG 12


RESULT   7
US-09-150-133-8
; Sequence 8, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
;  APPLICANT: The Board of Regents of the University of Oklahoma
;  TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
;  TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERAES, AND METHODS OF USE THEREOF
;  FILE REFERENCE: 5820.504
;  CURRENT APPLICATION NUMBER: US/09/150,133B
;  CURRENT FILING DATE: 1998-09-09
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 8
;   LENGTH: 1760
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-150-133-8


  Query Match             76.7%;  Score 13.8;  DB 3;  Length 1760;
  Best Local Similarity   88.2%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cacagtcgtacaccacg 18
          ||||||| | |||||||
Db    939 cacagtcctgcaccacg 955


RESULT   8
US-09-150-141-8
; Sequence 8, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
;  APPLICANT: The Board of Regents of the University of Oklahoma
;  TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
;  TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERAES, AND METHODS OF USE THEREOF
;  FILE REFERENCE: 5820.495
;  CURRENT APPLICATION NUMBER: US/09/150,141B
;  CURRENT FILING DATE: 1998-09-09
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 8
;   LENGTH: 1760
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-150-141-8


  Query Match             76.7%;  Score 13.8;  DB 3;  Length 1760;
  Best Local Similarity   88.2%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cacagtcgtacaccacg 18
          ||||||| | |||||||
Db    939 cacagtcctgcaccacg 955


RESULT   9
US-09-374-493-8
; Sequence 8, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
```

```
;  APPLICANT: The Board of Regents of the University of Oklahoma
;  TITLE OF INVENTION: TYROSYLPROETIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 5820.546
;  CURRENT APPLICATION NUMBER: US/09/374,493
;  CURRENT FILING DATE: 1999-08-13
;  EARLIER APPLICATION NUMBER: 09/150,133
;  EARLIER FILING DATE: 1998-09-09
;  EARLIER APPLICATION NUMBER: 60/072,994
;  EARLIER FILING DATE: 1998-01-29
;  EARLIER APPLICATION NUMBER: PCT/US99/16750
;  EARLIER FILING DATE: 1999-07-23
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 8
;   LENGTH: 1760
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-374-493-8


  Query Match             76.7%;  Score 13.8;  DB 4;  Length 1760;
  Best Local Similarity   88.2%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cacagtcgtacaccacg 18
          ||||||| | |||||||
Db    939 cacagtcctgcaccacg 955


RESULT  10
US-09-374-824-8
; Sequence 8, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
;  APPLICANT: The Board of Regents of the University of Oklahoma
;  TITLE OF INVENTION: TYROSYLPROETIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 5820.547
;  CURRENT APPLICATION NUMBER: US/09/374,824
;  CURRENT FILING DATE: 1999-08-13
;  EARLIER APPLICATION NUMBER: 09/150,133
;  EARLIER FILING DATE: 1998-09-09
;  EARLIER APPLICATION NUMBER: 60/072,994
;  EARLIER FILING DATE: 1998-01-29
;  EARLIER APPLICATION NUMBER: PCT/US99/16750
;  EARLIER FILING DATE: 1999-07-23
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 8
;   LENGTH: 1760
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-374-824-8


  Query Match             76.7%;  Score 13.8;  DB 4;  Length 1760;
  Best Local Similarity   88.2%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cacagtcgtacaccacg 18
          ||||||| | |||||||
Db    939 cacagtcctgcaccacg 955


RESULT  11
US-09-374-492-8
; Sequence 8, Application US/09374492
; Patent No. 6207432
; GENERAL INFORMATION:
;  APPLICANT: The Board of Regents of the University of Oklahoma
;  TITLE OF INVENTION: TYROSYLPROETIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 5820.545
;  CURRENT APPLICATION NUMBER: US/09/374,492
;  CURRENT FILING DATE: 1999-08-13
;  EARLIER APPLICATION NUMBER: 09/150,141
;  EARLIER FILING DATE: 1998-09-09
;  EARLIER APPLICATION NUMBER: 60/072,994
;  EARLIER FILING DATE: 1998-01-29
;  EARLIER APPLICATION NUMBER: PCT/US99/16750
;  EARLIER FILING DATE: 1999-07-23
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 8
;   LENGTH: 1760
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-374-492-8


  Query Match             76.7%;  Score 13.8;  DB 4;  Length 1760;
  Best Local Similarity   88.2%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cacagtcgtacaccacg 18
          ||||||| | |||||||
Db    939 cacagtcctgcaccacg 955


RESULT  12
US-08-891-298-2
; Sequence 2, Application US/08891298
; Patent No. 6300488
;  GENERAL INFORMATION:
;    APPLICANT:  Gage, Frederick H.
;    APPLICANT:  Suhr, Steven T.
;    TITLE OF INVENTION:  Modified Lepidopteran Receptors
;    TITLE OF INVENTION:  and Hybrid Multi-Functional Proteins for Use in Transcript
;    TITLE OF INVENTION:  and Transgene Expression Regulation
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gray Cary Ware & Freidenrich
;      STREET:  4365 Executive Drive, Suite 1600
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows DEMONSTRATION Version 2.0D
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/891,298
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-677-1409
;      TELEFAX:  619-677-1465
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2711 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  185...2002
;      OTHER INFORMATION:
US-08-891-298-2
```

```
  Query Match           76.7%;  Score 13.8;  DB 4;  Length 2711;
  Best Local Similarity  88.2%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccac 17
          ||||||||| | |||||
Db   1092 CCACAGTCGAAGACCAC 1108


RESULT  13
US-08-042-747A-4/c
; Sequence 4, Application US/08042747A
; Patent No. 5487969
;  GENERAL INFORMATION:
;    APPLICANT:  Eberle, Richard
;    APPLICANT:  Black, Darla
;    APPLICANT:  Scinicariello, Franco
;    APPLICANT:  Hilliard, Julia K.
;    TITLE OF INVENTION:  Cloning and Amplification of Monkey B
;    TITLE OF INVENTION:  Virus Genes
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cox & Smith Incorporated
;      STREET:  112 East Pecan Street, Suite 2000
;      CITY:  San Antonio
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  78205
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/042,747A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haymond, W. Bradley
;      REGISTRATION NUMBER:  35186
;      REFERENCE/DOCKET NUMBER:  S-0072.179
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  210-554-5500
;      TELEFAX:  210-226-8395
;      TELEX:  767609
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3177 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  269..2941
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..249
US-08-042-747A-4


  Query Match           76.7%;  Score 13.8;  DB 1;  Length 3177;
  Best Local Similarity  88.2%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cacagtcgtacaccacg 18
          | | ||||||||||||||
Db    166 CGCGGTCGTACACCACG 150


RESULT  14
US-08-755-587-27
; Sequence 27, Application US/08755587
; Patent No. 6045997
;  GENERAL INFORMATION:
;    APPLICANT:  Futreal, Phillip A
;    APPLICANT:  Wooster, Richard F
;    APPLICANT:  Ashworth, Alan
;    APPLICANT:  Stratton, Michael R
;    TITLE OF INVENTION:  Materials and methods relating to the
;    TITLE OF INVENTION:  identification and sequencing of the BRCA2 cancer
;    TITLE OF INVENTION:  susceptibility gene and uses thereof.
;    NUMBER OF SEQUENCES:  222
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson
;      STREET:  310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
;      CITY:  Raleigh
;      STATE:  NC
;      COUNTRY:  USA
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/755,587
;      FILING DATE:  25-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9523959.6
;      FILING DATE:  23-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9525555.0
;      FILING DATE:  14-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9617961.9
;      FILING DATE:  28-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kenneth D Sibley
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5405-135
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5892 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  483..5412
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  481..5412
US-08-755-587-27


  Query Match           76.7%;  Score 13.8;  DB 3;  Length 5892;
  Best Local Similarity  88.2%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccac 17
          |||||| | ||||||||
Db    235 CCACAGGCATACACCAC 251


RESULT  15
US-08-480-784-20/c
; Sequence 20, Application US/08480784
; Patent No. 5693473
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
```

```
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,784
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6769 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-480-784-20


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 6769;
  Best Local Similarity   88.2%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccac 17
           | |||| ||||||||||
Db     965 CTACAGGCGTACACCAC 949



Search completed: February 24, 2002, 03:01:32
Job time: 21443 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 21:47:34 ; Search time 6170.56 Seconds
(without alignments)
5781.645 Million cell updates/sec

Title: US-09-171-553B-1
Perfect score: 3320
Sequence: 1 gaattcgcggccgcgtcgac.....aaaaaaaaaaaaaaaaaaaa 3320

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	74	2.2	391	11	BF709087	BF709087 MI-P-AY0-
c	2	74	2.2	545	10	BE013835	BE013835 125326 MA
	3	57	1.7	399	11	BI183551	BI183551 UNL-P-FN-
	4	52	1.6	403	10	AW447645	AW447645 89634 MAR
c	5	52	1.6	451	13	AZ891915	AZ891915 RPCI-24-1
	6	50	1.5	549	13	AZ380763	AZ380763 1M0136B22
	7	42	1.3	576	13	AZ730616	AZ730616 RPCI-24-1
	8	40	1.2	704	13	BH059214	BH059214 RPCI-24-3
	9	39	1.2	100	11	BF442561	BF442561 259390 MA
	10	37	1.1	482	13	AQ920035	AQ920035 RPCI-23-2
	11	36	1.1	545	10	AA529580	AA529580 vi41d04.r
c	12	36	1.1	575	13	AZ479717	AZ479717 1M0300M05
c	13	36	1.1	583	13	AZ415280	AZ415280 1M0190M06
	14	36	1.1	599	13	AZ944719	AZ944719 2M0205L23
	15	36	1.1	610	13	AZ859488	AZ859488 2M0165D07
c	16	36	1.1	640	13	AZ997740	AZ997740 2M0284F23
	17	36	1.1	646	13	AZ638811	AZ638811 1M0498I23
	18	36	1.1	655	13	AZ740414	AZ740414 RPCI-24-1
c	19	36	1.1	660	13	AZ353293	AZ353293 1M0092E07
c	20	36	1.1	673	13	AZ329067	AZ329067 1M0053G10
	21	34	1.0	345	13	AZ780588	AZ780588 2M0018A11
c	22	34	1.0	406	13	AZ506946	AZ506946 1M0348P14
	23	34	1.0	550	13	AZ390750	AZ390750 1M0152P16
	24	34	1.0	574	13	AZ390749	AZ390749 1M0152P15
c	25	34	1.0	630	13	AZ838774	AZ838774 2M0134D13
c	26	33	1.0	396	10	AA736984	AA736984 nx89a09.s
	27	33	1.0	402	11	BG671688	BG671688 DRNBWE08
	28	33	1.0	455	13	AZ818850	AZ818850 2M0089D10
	29	32	1.0	206	13	AZ792975	AZ792975 2M0046F05
	30	32	1.0	229	13	AZ914781	AZ914781 RPCI-24-1
	31	32	1.0	298	10	AW102270	AW102270 sd85e09.y
c	32	32	1.0	353	10	AA918868	AA918868 ol70a12.s
	33	32	1.0	358	11	BF604441	BF604441 270277 MA
	34	32	1.0	361	10	AW074793	AW074793 xa70e06.x
c	35	32	1.0	465	13	AZ431949	AZ431949 1M0217I22
c	36	32	1.0	502	13	AZ447630	AZ447630 1M0244F20
c	37	32	1.0	512	13	AZ459290	AZ459290 1M0264D05
c	38	32	1.0	518	10	AL499730	AL499730 AL499730
c	39	32	1.0	523	13	AZ995885	AZ995885 2M0281E19
	40	32	1.0	556	11	BG606618	BG606618 WHE2958_H
	41	32	1.0	561	13	AZ793135	AZ793135 2M0046F07
	42	32	1.0	561	13	AZ951769	AZ951769 2M0216O15
c	43	32	1.0	567	13	AZ624307	AZ624307 1M0462K20
	44	32	1.0	568	13	AZ824214	AZ824214 2M0098N07
c	45	32	1.0	573	13	AZ797677	AZ797677 2M0054G04

ALIGNMENTS

RESULT 1
BF709087/c
LOCUS BF709087 391 bp mRNA EST 02-JAN-2001
DEFINITION MI-P-AY0-nae-c-07-0-UI.s1 MI-P-AY0 Sus scrofa cDNA clone MI-P-AY0-nae-c-07-0-UI 3', mRNA sequence.
ACCESSION BF709087
VERSION BF709087.1 GI:12008564
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 391)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized placenta library cDNA Library Preparation: M.B. Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab, University of Iowa Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward

```
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .391
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone="MI-P-AY0-nae-c-07-0-UI"
                     /clone_lib="MI-P-AY0"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-AY0
                     library is derived from placenta. For a detailed
                     description of the  library from which this clone was
                     derived, please visit our web site at
                     http://pigest.genome.iastate.edu/. The procedure used to
                     create this library has been previously described (Bonaldo
                     , Lennon and Soares, Genome Research 6:791-806, 1996)
                     TAG_LIB=MI-P-AY0
                     TAG_TISSUE=placenta
                     TAG_SEQ=ATTGG"
BASE COUNT       94 a     87 c     71 g    139 t
ORIGIN


  Query Match             2.2%;  Score 74;  DB 11;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-11;
  Matches   74;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1984 gtttttgtagacacctttcaggatgggtagaggcttatcctactaagaaagagacttca 2043
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 GTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCA 191

Qy   2044 accgtggtggctaa 2057
           ||||||||||||||
Db    190 ACCGTGGTGGCTAA 177


RESULT   2
BE013835/c
LOCUS       BE013835      545 bp    mRNA            EST       09-JUL-2000
DEFINITION  125326 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BE013835
VERSION     BE013835.1  GI:8274822
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 52  row: B  column: 23
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT      128 a    120 c    129 g    168 t
ORIGIN


  Query Match             2.2%;  Score 74;  DB 10;  Length 545;
  Best Local Similarity   100.0%;  Pred. No. 2e-11;
  Matches   74;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1984 gtttttgtagacacctttcaggatgggtagaggcttatcctactaagaaagagacttca 2043
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    216 GTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCA 157

Qy   2044 accgtggtggctaa 2057
           ||||||||||||||
Db    156 ACCGTGGTGGCTAA 143


RESULT   3
BI183551
LOCUS       BI183551      399 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-by-g-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-by-g-02-0-UNL 3', mRNA sequence.
ACCESSION   BI183551
VERSION     BI183551.1  GI:14657960
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .399
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-by-g-02-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT      120 a     83 c     92 g     96 t      8 others
```

```
ORIGIN


  Query Match             1.7%;  Score 57;  DB 11;  Length 399;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-06;
  Matches   57;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1920 cgctcactgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaata 1976
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    116 CGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAATACGGAAACAAATA 172


RESULT   4
AW447645
LOCUS       AW447645      403 bp    mRNA            EST       25-APR-2001
DEFINITION  89634 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW447645
VERSION     AW447645.1  GI:6989432
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 70  row: L  column: 19
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .403
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT       61 a    134 c    124 g     84 t
ORIGIN


  Query Match             1.6%;  Score 52;  DB 10;  Length 403;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-05;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2461 gagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggag 2512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 GAGTGGGTGAGGCAGCGAGCGTGGAAGCAGCTCCGGGAGGCCTACTCAGGAG 353


RESULT   5
AZ891915/c
LOCUS       AZ891915      451 bp    DNA             GSS       05-MAR-2001
DEFINITION  RPCI-24-172A20.TV RPCI-24 Mus musculus genomic clone RPCI-24-172A20
            , DNA sequence.
ACCESSION   AZ891915
VERSION     AZ891915.1  GI:13210860
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-172A20.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 172  row: A  column: 20
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .451
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-172A20"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      144 a     91 c    106 g    110 t
ORIGIN


  Query Match             1.6%;  Score 52;  DB 13;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-05;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2420 tgctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgag 2471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 TGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAG 253


RESULT   6
AZ380763
LOCUS       AZ380763      549 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0136B22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0136B22 R, DNA sequence.
ACCESSION   AZ380763
VERSION     AZ380763.1  GI:10494463
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
```

```
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0136  row: B  column: 22
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 549.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0136B22"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      159 a    119 c    119 g    152 t
ORIGIN


  Query Match             1.5%;  Score 50;  DB 13;  Length 549;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-05;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2131 gttgcccaggtaagtcagggactggccaagatattggggattgattggaa 2180
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    228 GTTGCCCAGGTAAGTCAGGGACTGGCCAAGATATTGGGGATTGATTGGAA 277


RESULT   7
AZ730616
LOCUS       AZ730616      576 bp    DNA             GSS       25-JAN-2001
DEFINITION  RPCI-24-149D1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-149D1,
            DNA sequence.
ACCESSION   AZ730616
VERSION     AZ730616.1  GI:12491859
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-149D1.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 149  row: D  column: 1
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-149D1"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      161 a    139 c    121 g    155 t
ORIGIN


  Query Match             1.3%;  Score 42;  DB 13;  Length 576;
  Best Local Similarity   100.0%;  Pred. No. 0.016;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2406 acatagtgctgatgtgctgctttcccagcctttgttctctag 2447
           |||||||||||||||||||||||||||||||||||||||||
Db    317 ACATAGTGCTGATGTGCTGCTTTCCCAGCCTTTGTTCTCTAG 358


RESULT   8
BH059214
LOCUS       BH059214      704 bp    DNA             GSS       18-JUL-2001
DEFINITION  RPCI-24-326H3.TJ RPCI-24 Mus musculus genomic clone RPCI-24-326H3,
            DNA sequence.
ACCESSION   BH059214
VERSION     BH059214.1  GI:14868999
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 704)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-326H3.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
```

```
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 326  row: H  column: 3
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .704
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-326H3"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      224 a    132 c    174 g    174 t
ORIGIN


  Query Match             1.2%;  Score 40;  DB 13;  Length 704;
  Best Local Similarity   100.0%;  Pred. No. 0.049;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2137 caggtaagtcagggactggccaagatattggggattgatt 2176
           |||||||||||||||||||||||||||||||||||||||||
Db     562 CAGGTAAGTCAGGGACTGGCCAAGATATTGGGGATTGATT 601


RESULT   9
BF442561
LOCUS       BF442561      100 bp    mRNA            EST       01-DEC-2000
DEFINITION  259390 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF442561
VERSION     BF442561.1  GI:11502653
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 100)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 90  row: I  column: 16
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .100
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT       29 a     24 c     26 g     21 t
ORIGIN


  Query Match             1.2%;  Score 39;  DB 11;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1348 ggggccatctataaacaaaggggttgcttacctcagca 1386
           |||||||||||||||||||||||||||||||||||||||
Db      55 GGGGCCATCTATAAACAAAGGGGGTTGCTTACCTCAGCA 93


RESULT  10
AQ920035
LOCUS       AQ920035      482 bp    DNA             GSS       21-DEC-1999
DEFINITION  RPCI-23-273D7.TV RPCI-23 Mus musculus genomic clone RPCI-23-273D7,
            DNA sequence.
ACCESSION   AQ920035
VERSION     AQ920035.1  GI:6609038
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 273  row: D  column: 7
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-273D7"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      128 a     91 c    106 g    154 t      3 others
ORIGIN


  Query Match             1.1%;  Score 37;  DB 13;  Length 482;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2420 tgctgctttcccagcctttgttctctaggctcaaggc 2456
          |||||||||||||||||||||||||||||||||||||
Db     29 TGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGC 65


RESULT  11
AA529580
LOCUS       AA529580      545 bp    mRNA            EST       22-JUL-1997
DEFINITION  vi41d04.r1 Beddington mouse embryonic region Mus musculus cDNA
            clone IMAGE:906343 5' similar to SW:POL_GALV P21414 POL POLYPROTEIN
            ;, mRNA sequence.
ACCESSION   AA529580
VERSION     AA529580.1  GI:2272286
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:527007
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 326.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Mus musculus"
                     /strain="C57BL6 x DBA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:906343"
                     /clone_lib="Beddington mouse embryonic region"
                     /sex="pooled"
                     /tissue_type="embryo"
                     /dev_stage="7.5dpc"
                     /lab_host="DH12S"
                     /note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT. Gastrulating embryos were collected at 7.5dpc
                     from C57BL6 x DBA matings, excluding embryos that had
                     developed head folds and all extraembryonic tissues.
                     Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
                     Referenced in Development 121, 2479-2489 (1995)"
BASE COUNT      143 a    130 c    127 g    145 t
ORIGIN


  Query Match             1.1%;  Score 36;  DB 10;  Length 545;
  Best Local Similarity   100.0%;  Pred. No. 0.78;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2420 tgctgctttcccagcctttgttctctaggctcaagg 2455
          |||||||||||||||||||||||||||||||||||
Db    176 TGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGG 211


RESULT  12
AZ479717/c
LOCUS       AZ479717      575 bp    DNA             GSS       04-OCT-2000
DEFINITION  1M0300M05R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0300M05 R, DNA sequence.
ACCESSION   AZ479717
VERSION     AZ479717.1  GI:10639886
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0300  row: M  column: 05
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 575.
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0300M05"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      136 a    140 c    137 g    162 t
ORIGIN


  Query Match             1.1%;  Score 36;  DB 13;  Length 575;
  Best Local Similarity   100.0%;  Pred. No. 0.75;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2203 caaagctcaggacaggtagagaggatgaatagaacc 2238
          |||||||||||||||||||||||||||||||||||
Db    160 CAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACC 125
```

```
RESULT  13
AZ415280/c
LOCUS       AZ415280      583 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0190M06F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0190M06 F, DNA sequence.
ACCESSION   AZ415280
VERSION     AZ415280.1  GI:10539293
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0190  row: M  column: 06
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 583.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0190M06"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      146 a    149 c    130 g    158 t
ORIGIN


  Query Match             1.1%;  Score 36;  DB 13;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 0.74;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2203 caaagctcaggacaggtagagaggatgaatagaacc 2238
           ||||||||||||||||||||||||||||||||||||
Db     193 CAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACC 158
```

```
RESULT  14
AZ944719
LOCUS       AZ944719      599 bp    DNA             GSS       26-APR-2001
DEFINITION  2M0205L23R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0205L23 R, DNA sequence.
ACCESSION   AZ944719
VERSION     AZ944719.1  GI:13809008
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0205  row: L  column: 23
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 599.
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0205L23"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      160 a    136 c    151 g    152 t
ORIGIN


  Query Match             1.1%;  Score 36;  DB 13;  Length 599;
  Best Local Similarity   100.0%;  Pred. No. 0.72;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2203 caaagctcaggacaggtagagaggatgaatagaacc 2238
           ||||||||||||||||||||||||||||||||||||
Db     334 CAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACC 369
```

```
RESULT 15
AZ859488
LOCUS       AZ859488      610 bp    DNA             GSS       21-FEB-2001
DEFINITION  2M0165D07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0165D07 F, DNA sequence.
ACCESSION   AZ859488
VERSION     AZ859488.1  GI:13053858
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0165  row: D  column: 07
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 610.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0165D07"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      194 a    145 c    130 g    141 t
ORIGIN


  Query Match          1.1%;  Score 36;  DB 13;  Length 610;
  Best Local Similarity   100.0%;  Pred. No. 0.71;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2420 tgctgctttcccagcctttgttctctaggctcaagg 2455
           |||||||||||||||||||||||||||||||||||
Db    410 TGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGG 445


Search completed: February 24, 2002, 04:28:16
Job time: 24042 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 04:28:16 ; Search time 6170.56 Seconds
(without alignments)
14295.639 Million cell updates/sec

Title: US-09-171-553B-3
Perfect score: 8209
Sequence: 1 gtggtgtacgactgtgggcc.....aaaaaaaaaaaaaaaaaaaa 8209

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	190	2.3	564	11	BI182930	BI182930 UNL-P-FN-
	2	140	1.7	466	11	BF441468	BF441468 257862 MA
	3	140	1.7	467	11	BF441465	BF441465 257856 MA
c	4	130	1.6	532	11	BF709993	BF709993 MI-P-AY1-
c	5	126	1.5	408	11	BF711018	BF711018 MI-P-AY1-
	6	102	1.2	564	11	BI181099	BI181099 UNL-P-FN-
	7	102	1.2	564	11	BI186129	BI186129 UNL-P-FN-
	8	96	1.2	376	11	BF441469	BF441469 257863 MA
	9	92	1.1	558	11	BI360879	BI360879 388957 MA
c	10	92	1.1	602	11	BF712162	BF712162 MI-P-E6-a
	11	80	1.0	452	11	BF441466	BF441466 257857 MA
	12	78	1.0	549	10	AW657531	AW657531 110922 MA
	13	77	0.9	857	11	BI184013	BI184013 UNL-P-FN-
c	14	74	0.9	391	11	BF709087	BF709087 MI-P-AY0-
c	15	74	0.9	545	10	BE013835	BE013835 125326 MA
	16	72	0.9	229	11	BF443400	BF443400 260872 MA
c	17	69	0.8	468	11	BI186066	BI186066 UNL-P-FN-
c	18	69	0.8	499	11	BF704354	BF704354 MI-P-O3-a
c	19	69	0.8	521	11	BI399852	BI399852 MI-P-AY1-
	20	69	0.8	730	11	BI185535	BI185535 UNL-P-FN-
c	21	68	0.8	468	11	BI182727	BI182727 UNL-P-FN-
c	22	64	0.8	447	11	BI401063	BI401063 MI-P-CP0-
c	23	64	0.8	485	11	BF703671	BF703671 MI-P-E4-a
c	24	64	0.8	554	11	BF703805	BF703805 MI-P-O2-a
	25	63	0.8	575	11	BI183356	BI183356 UNL-P-FN-
c	26	59	0.7	465	11	BI399234	BI399234 MI-P-AY1-
	27	57	0.7	399	11	BI183551	BI183551 UNL-P-FN-
c	28	55	0.7	481	10	AW231947	AW231947 2C22N1 Su
	29	52	0.6	403	10	AW447645	AW447645 89634 MAR
c	30	52	0.6	451	13	AZ891915	AZ891915 RPCI-24-1
c	31	51	0.6	545	11	BF701947	BF701947 MI-P-E3-a
c	32	50	0.6	494	11	BI182742	BI182742 UNL-P-FN-
	33	50	0.6	549	13	AZ380763	AZ380763 1M0136B22
c	34	49	0.6	300	11	BI184279	BI184279 UNL-P-FN-
	35	48	0.6	345	10	AW358862	AW358862 43799 MAR
	36	48	0.6	595	11	BI185465	BI185465 UNL-P-FN-
c	37	47	0.6	294	11	BF713588	BF713588 MI-P-O2-a
c	38	47	0.6	579	11	BI183723	BI183723 UNL-P-FN-
c	39	45	0.5	286	13	AZ771375	AZ771375 1M0573C02
	40	45	0.5	575	13	BH025657	BH025657 RPCI-24-3
c	41	45	0.5	599	13	AZ041781	AZ041781 RPCI-23-2
	42	44	0.5	173	10	AW435835	AW435835 75099 MAR
c	43	44	0.5	341	10	AW416859	AW416859 52601 MAR
	44	42	0.5	576	13	AZ730616	AZ730616 RPCI-24-1
c	45	41	0.5	256	10	AW308385	AW308385 3994 MARC

ALIGNMENTS

RESULT 1
BI182930/c
LOCUS BI182930 564 bp mRNA EST 10-JUL-2001
DEFINITION UNL-P-FN-br-h-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone UNL-P-FN-br-h-08-0-UNL 3', mRNA sequence.
ACCESSION BI182930
VERSION BI182930.1 GI:14657339
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 564)
AUTHORS Caetano,A.R., Johnson,R.K. and Pomp,D.
TITLE Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
JOURNAL Unpublished (2001)
COMMENT Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized porcine ovarian follicles library
Seq primer: M13 -29
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .564
/organism="Sus scrofa"

```
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-br-h-08-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_LIB=UNL-P-FN
                     TAG_TISSUE=porcine ovarian follicles
                     TAG_SEQ=CACACT"
BASE COUNT      141 a    129 c    143 g    151 t
ORIGIN


  Query Match             2.3%;  Score 190;  DB 11;  Length 564;
  Best Local Similarity   99.3%;  Pred. No. 9.3e-39;
  Matches  290;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   7720 cctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggcctgcta 7779
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    564 CCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTA 505

Qy   7780 tcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggattctgtaa 7839
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    504 TCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAA 445

Qy   7840 aactgactggcaccatagaagaattgattacacattgacagccctagtgacctatctcaa 7899
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    444 AACTGACTGGCACCATAGAAGAACTGATTACACATTGACAGCCCTAGTGACCTATCTCAA 385

Qy   7900 ctgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagctattttaa 7959
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    384 CTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGACCTCCAGAGCTATTTTAA 325

Qy   7960 aatgattggtccacggagcgcgggctctcgatattttaaaatgattggtcca 8011
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    324 AATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCA 273


RESULT   2
BF441468
LOCUS       BF441468      466 bp    mRNA            EST       01-DEC-2000
DEFINITION  257862 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF441468
VERSION     BF441468.1  GI:11501632
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 86  row: D  column: 8
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      114 a    118 c    127 g    107 t
ORIGIN


  Query Match             1.7%;  Score 140;  DB 11;  Length 466;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-26;
  Matches  140;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1359 cagctgctgcagacactcttcacaaccgaggagcgagagagaattctgttagaggctaga 1418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGCTGCTGCAGACACTCTTCACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAGA 120

Qy   1419 aaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatgagattgacatggga 1478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAAAATGTTCCTGGGGCCGACGGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGA 180

Qy   1479 tttcccttgactcgccccgg 1498
           ||||||||||||||||||||
Db    181 TTTCCCTTGACTCGCCCCGG 200


RESULT   3
BF441465
LOCUS       BF441465      467 bp    mRNA            EST       01-DEC-2000
DEFINITION  257856 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF441465
VERSION     BF441465.1  GI:11501629
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 86  row: C  column: 8
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
```

```
     source          1. .467
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      115 a    118 c    127 g    107 t
ORIGIN


  Query Match             1.7%;  Score 140;  DB 11;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-26;
  Matches  140;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1359 cagctgctgcagacactcttcacaaccgaggagcgagagagaattctgttagaggctaga 1418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGCTGCTGCAGACACTCTTCACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAGA 120

Qy   1419 aaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatgagattgacatggga 1478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAAAATGTTCCTGGGGCCGACGGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGA 180

Qy   1479 tttcccttgactcgccccgg 1498
          ||||||||||||||||||||
Db    181 TTTCCCTTGACTCGCCCCGG 200


RESULT   4
BF709993/c
LOCUS       BF709993      532 bp    mRNA            EST       02-JAN-2001
DEFINITION  MI-P-AY1-nqa-b-06-0-UI.s1 MI-P-AY1 Sus scrofa cDNA clone
            MI-P-AY1-nqa-b-06-0-UI 3', mRNA sequence.
ACCESSION   BF709993
VERSION     BF709993.1  GI:12009470
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,
            University of Iowa Clone distribution: clones will be available
            through Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 59-132,
            >GC_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-AY1-nqa-b-06-0-UI"
                     /clone_lib="MI-P-AY1"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-AY1
                     library is normalized library derived from the MI-P-AY0
                     library,  ultimately derived from placenta tissue. For a
                     detailed description of the  library from which this clone
                     was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/. The procedure used to
                     create this library has been previously described (Bonaldo
                     , Lennon and Soares, Genome Research 6: 791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT      124 a    184 c    154 g     70 t
ORIGIN


  Query Match             1.6%;  Score 130;  DB 11;  Length 532;
  Best Local Similarity   99.4%;  Pred. No. 1.3e-23;
  Matches  180;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    364 ccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctactgtc 423
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    427 CCCTGGGGGACGCCCCGGGAGGTGGGGAGAGCCAGGGACGCCTGGTGGTCTCCTTCTGTC 368

Qy    424 ggtcagaggaccgagttctgttgttgaagcgaaagcttccccctccgcggccgtccgact 483
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367 GGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCCGACT 308

Qy    484 cttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggtttctg 543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 CTTTTGCCTGCTTGTGGAAGACGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGTTTCTG 248

Qy    544 t 544
          |
Db    247 T 247


RESULT   5
BF711018/c
LOCUS       BF711018      408 bp    mRNA            EST       02-JAN-2001
DEFINITION  MI-P-AY1-nrb-h-02-0-UI.s1 MI-P-AY1 Sus scrofa cDNA clone
            MI-P-AY1-nrb-h-02-0-UI 3', mRNA sequence.
ACCESSION   BF711018
VERSION     BF711018.1  GI:12010495
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,
            University of Iowa Clone distribution: clones will be available
            through Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 59-132,
            >GC_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
```

```
                     /clone="MI-P-AY1-nrb-h-02-0-UI"
                     /clone_lib="MI-P-AY1"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-AY1
                     library is normalized library derived from the MI-P-AY0
                     library,  ultimately derived from placenta tissue. For a
                     detailed description of the  library from which this clone
                     was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/. The procedure used to
                     create this library has been previously described (Bonaldo
                     , Lennon and Soares, Genome Research 6: 791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT       94 a    139 c    122 g     52 t      1 others
ORIGIN


  Query Match             1.5%;  Score 126;  DB 11;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-22;
  Matches  126;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     419 ctgtcggtcagaggaccgagttctgttgttgaagcgaaagcttccccctccgcggccgtc 478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     372 CTGTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTC 313

Qy     479 cgactcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggt 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     312 CGACTCTTTTGCCTGCTTGTGGAAGACGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGT 253

Qy     539 ttctgt 544
           ||||||
Db     252 TTCTGT 247


RESULT   6
BI181099
LOCUS       BI181099      564 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-ab-d-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-ab-d-02-0-UNL 3', mRNA sequence.
ACCESSION   BI181099
VERSION     BI181099.1  GI:14655508
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-ab-d-02-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT      112 a    196 c    143 g    112 t      1 others
ORIGIN


  Query Match             1.2%;  Score 102;  DB 11;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-16;
  Matches  102;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1204 gccaattgcagcccctccagtattggccctttttcttctgcagatctctataattggaaaa 1263
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     372 GCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGGAAAA 431

Qy    1264 ctaaccatccccctttctcggaggatccccaacgcctcacgg 1305
           |||||||||||||||||||||||||||||||||||||||||||
Db     432 CTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGG 473


RESULT   7
BI186129
LOCUS       BI186129      564 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-ab-d-02-0-UNL.s2 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-ab-d-02-0-UNL 3', mRNA sequence.
ACCESSION   BI186129
VERSION     BI186129.1  GI:14660538
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-ab-d-02-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
```

```
                nucleotides present between the Not I site and the
                oligo-dT track.  The library was constructed as described
                by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                , 1996.
                TAG_SEQ=None found"
BASE COUNT      112 a    196 c    144 g    112 t
ORIGIN


  Query Match             1.2%;  Score 102;  DB 11;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-16;
  Matches  102;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1204 gccaattgcagccccctccagtattggccctttttcttctgcagatctctataattggaaaaa 1263
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 GCCAATTGCAGCCCCTCCAGTATTGGCCCTTTTCTTCTGCAGATCTCTATAATTGGAAAA 431

Qy   1264 ctaaccatccccctttctcggaggatccccaacgcctcacgg 1305
          |||||||||||||||||||||||||||||||||||||||||
Db    432 CTAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCACGG 473


RESULT   8
BF441469
LOCUS       BF441469      376 bp    mRNA            EST       01-DEC-2000
DEFINITION  257863 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF441469
VERSION     BF441469.1  GI:11501633
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 376)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 86  row: D  column: 9
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .376
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT       91 a     94 c    108 g     83 t
ORIGIN


  Query Match             1.2%;  Score 96;  DB 11;  Length 376;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-15;
  Matches   96;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1359 cagctgctgcagacactcttcacaaccgaggagcgagagagaattctgttagaggctaga 1418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGCTGCTGCAGACACTCTTCACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAGA 120

Qy   1419 aaaaatgttcctggggccgacgggcgacccacgcag 1454
          |||||||||||||||||||||||||||||||||||
Db    121 AAAAATGTTCCTGGGGCCGACGGGCGACCCACGCAG 156


RESULT   9
BI360879
LOCUS       BI360879      558 bp    mRNA            EST       01-AUG-2001
DEFINITION  388957 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BI360879
VERSION     BI360879.1  GI:15056907
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 142  row: B  column: 4
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      129 a    123 c    119 g    187 t
ORIGIN


  Query Match             1.1%;  Score 92;  DB 11;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-14;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   8098 gcactcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgct 8157
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 GCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCT 191

Qy   8158 tggaataaaaatcctcttgctgtttgcatcaa 8189
          |||||||||||||||||||||||||||||||
Db    192 TGGAATAAAAATCCTCTTGCTGTTTGCATCAA 223


RESULT  10
BF712162/c
LOCUS       BF712162      602 bp    mRNA            EST       02-JAN-2001
DEFINITION  MI-P-E6-acx-e-11-1-UM.s1 MI-P-E6 Sus scrofa cDNA clone
            MI-P-E6-acx-e-11-1-UM 3', mRNA sequence.
ACCESSION   BF712162
```

```
VERSION     BF712162.1  GI:12011595
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized fetus at gestational day 20 library cDNA Library
            Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
            Research Center, Department of Animal Science, University of
            Missouri-Columbia, 65211 Clone distribution: clones will be
            available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-E6-acx-e-11-1-UM"
                     /clone_lib="MI-P-E6"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-E6
                     library is derived from fetus at gestational day 20.  For
                     a detailed description of the  library from which this
                     clone was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/.
                     TAG_LIB=MI-P-E6
                     TAG_TISSUE=fetus at gestational day 20
                     TAG_SEQ=AGGAA"
BASE COUNT      141 a    146 c    152 g    162 t      1 others
ORIGIN


  Query Match             1.1%;  Score 92;  DB 11;  Length 602;
  Best Local Similarity   99.3%;  Pred. No. 4.3e-14;
  Matches  142;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    8060 ggctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgcagtcctct 8119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     143 GGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGTCCTCT 84

Qy    8120 acccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctg 8179
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      83 ACCCCTGCGTGGCGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTG 24

Qy    8180 tttgcatcaaaaaaaaaaaaaaa 8202
           |||||||||||||||||||||||
Db      23 TTTGCATCAAAAAAAAAAAAAAA 1


RESULT  11
BF441466
LOCUS       BF441466      452 bp    mRNA            EST       01-DEC-2000
DEFINITION  257857 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF441466
VERSION     BF441466.1  GI:11501630
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 86  row: C  column: 9
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      107 a    118 c    125 g    102 t
ORIGIN


  Query Match             1.0%;  Score 80;  DB 11;  Length 452;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-11;
  Matches   80;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1419 aaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatgagattgacatggga 1478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 AAAAATGTTCCTGGGGCCGACGGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGA 179

Qy    1479 tttcccttgactcgccccgg 1498
           ||||||||||||||||||||
Db     180 TTTCCCTTGACTCGCCCCGG 199


RESULT  12
AW657531
LOCUS       AW657531      549 bp    mRNA            EST       25-APR-2001
DEFINITION  110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW657531
VERSION     AW657531.1  GI:7423429
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
```

```
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 94  row: E  column: 12
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT      145 a    144 c    121 g    139 t
ORIGIN


  Query Match             1.0%;  Score 78;  DB 10;  Length 549;
  Best Local Similarity   99.2%;  Pred. No. 1.6e-10;
  Matches  128;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    6640 aaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaataagctt 6699
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159 AAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTC 218

Qy    6700 accctcactgaagtttccgggaaggggacatgcataggaaaagctcccccatcccaccaa 6759
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     219 ACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAA 278

Qy    6760 cacctttgc 6768
           |||||||||
Db     279 CACCTTTGC 287


RESULT  13
BI184013
LOCUS       BI184013      857 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-cf-b-12-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-cf-b-12-0-UNL 3', mRNA sequence.
ACCESSION   BI184013
VERSION     BI184013.1  GI:14658422
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 857)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .857
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-cf-b-12-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT      152 a    217 c    248 g    236 t      4 others
ORIGIN


  Query Match             0.9%;  Score 77;  DB 11;  Length 857;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-10;
  Matches   77;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      28 gcttggaataaaaatcctcttgctgtttgcatcaagaccgcttctcgtgagtgatttggg 87
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38 GCTTGGAATAAAAATCCTCTTGCTGTTTGCATCAAGACCGCTTCTCGTGAGTGATTTGGG 97

Qy      88 gtgtcgcctcttccgag 104
           |||||||||||||||||
Db      98 GTGTCGCCTCTTCCGAG 114


RESULT  14
BF709087/c
LOCUS       BF709087      391 bp    mRNA            EST       02-JAN-2001
DEFINITION  MI-P-AY0-nae-c-07-0-UI.s1 MI-P-AY0 Sus scrofa cDNA clone
            MI-P-AY0-nae-c-07-0-UI 3', mRNA sequence.
ACCESSION   BF709087
VERSION     BF709087.1  GI:12008564
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 391)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
```

```
           non-normalized placenta library cDNA Library Preparation: M.B.
           Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,
           University of Iowa Clone distribution: clones will be available
           through Research Genetics (www.resgen.com)
           Seq primer: M13 Forward
           POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .391
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone="MI-P-AY0-nae-c-07-0-UI"
                     /clone_lib="MI-P-AY0"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-AY0
                     library is derived from placenta. For a detailed
                     description of the  library from which this clone was
                     derived, please visit our web site at
                     http://pigest.genome.iastate.edu/. The procedure used to
                     create this library has been previously described (Bonaldo
                     , Lennon and Soares, Genome Research 6:791-806, 1996)
                     TAG_LIB=MI-P-AY0
                     TAG_TISSUE=placenta
                     TAG_SEQ=ATTGG"
BASE COUNT       94 a     87 c     71 g    139 t
ORIGIN


  Query Match              0.9%;  Score 74;  DB 11;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-09;
  Matches   74;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4935 gtttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagacttca 4994
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 GTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCA 191

Qy   4995 accgtggtggctaa 5008
           ||||||||||||||
Db    190 ACCGTGGTGGCTAA 177


RESULT  15
BE013835/c
LOCUS       BE013835      545 bp    mRNA            EST       09-JUL-2000
DEFINITION  125326 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BE013835
VERSION     BE013835.1  GI:8274822
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 52  row: B  column: 23
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT      128 a    120 c    129 g    168 t
ORIGIN


  Query Match              0.9%;  Score 74;  DB 10;  Length 545;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-09;
  Matches   74;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4935 gtttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagacttca 4994
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    216 GTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCA 157

Qy   4995 accgtggtggctaa 5008
           ||||||||||||||
Db    156 ACCGTGGTGGCTAA 143


Search completed: February 24, 2002, 04:28:43
Job time: 24069 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 04:28:43 ; Search time 6170.56 Seconds
(without alignments)
5147.753 Million cell updates/sec

Title: US-09-171-553B-9
Perfect score: 2956
Sequence: 1 tgctttttagggttaggaac.....aaaaaaaaaaaaaaaaaaaa 2956

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	190	6.4	564	11	BI182930	BI182930 UNL-P-FN-
2	92	3.1	558	11	BI360879	BI360879 388957 MA
c 3	92	3.1	602	11	BF712162	BF712162 MI-P-E6-a
4	78	2.6	549	10	AW657531	AW657531 110922 MA
c 5	69	2.3	468	11	BI186066	BI186066 UNL-P-FN-
c 6	69	2.3	499	11	BF704354	BF704354 MI-P-O3-a
c 7	69	2.3	521	11	BI399852	BI399852 MI-P-AY1-
8	69	2.3	730	11	BI185535	BI185535 UNL-P-FN-
c 9	68	2.3	468	11	BI182727	BI182727 UNL-P-FN-
c 10	64	2.2	485	11	BF703671	BF703671 MI-P-E4-a
c 11	64	2.2	554	11	BF703805	BF703805 MI-P-O2-a
c 12	52	1.8	451	13	AZ891915	AZ891915 RPCI-24-1
c 13	51	1.7	545	11	BF701947	BF701947 MI-P-E3-a
c 14	49	1.7	300	11	BI184279	BI184279 UNL-P-FN-
c 15	48	1.6	494	11	BI182742	BI182742 UNL-P-FN-
16	48	1.6	595	11	BI185465	BI185465 UNL-P-FN-
c 17	47	1.6	294	11	BF713588	BF713588 MI-P-O2-a
18	47	1.6	576	13	AZ730616	AZ730616 RPCI-24-1
c 19	47	1.6	579	11	BI183723	BI183723 UNL-P-FN-
20	40	1.4	166	10	AW346655	AW346655 29171 MAR
c 21	40	1.4	631	11	BF702741	BF702741 MI-P-E3-a
22	38	1.3	403	10	AW447645	AW447645 89634 MAR
23	37	1.3	482	13	AQ920035	AQ920035 RPCI-23-2
24	36	1.2	545	10	AA529580	AA529580 vi41d04.r
25	36	1.2	610	13	AZ859488	AZ859488 2M0165D07
c 26	36	1.2	640	13	AZ997740	AZ997740 2M0284F23
27	36	1.2	646	13	AZ638811	AZ638811 1M0498I23
28	36	1.2	655	13	AZ740414	AZ740414 RPCI-24-1
c 29	36	1.2	660	13	AZ353293	AZ353293 1M0092E07
c 30	36	1.2	673	13	AZ329067	AZ329067 1M0053G10
31	35	1.2	857	11	BI184013	BI184013 UNL-P-FN-
32	34	1.2	155	10	AW346676	AW346676 29231 MAR
c 33	34	1.2	261	11	BI184980	BI184980 UNL-P-FN-
34	34	1.2	345	13	AZ780588	AZ780588 2M0018A11
c 35	34	1.2	513	10	AW786160	AW786160 119054 MA
36	34	1.2	550	13	AZ390750	AZ390750 1M0152P16
37	34	1.2	574	13	AZ390749	AZ390749 1M0152P15
c 38	32	1.1	340	13	BH068193	BH068193 RPCI-24-3
c 39	32	1.1	348	13	AZ966390	AZ966390 2M0236A21
40	31	1.0	251	11	BG382013	BG382013 297673 MA
41	31	1.0	252	10	AA245675	AA245675 mx05g11.r
42	31	1.0	437	11	BI081943	BI081943 602877172
43	31	1.0	834	13	BH124666	BH124666 RPCI-24-2
c 44	30	1.0	114	11	BF728989	BF728989 1000070E0
45	30	1.0	230	10	AW786010	AW786010 117582 MA

ALIGNMENTS

RESULT 1
BI182930/c
LOCUS BI182930 564 bp mRNA EST 10-JUL-2001
DEFINITION UNL-P-FN-br-h-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone UNL-P-FN-br-h-08-0-UNL 3', mRNA sequence.
ACCESSION BI182930
VERSION BI182930.1 GI:14657339
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 564)
AUTHORS Caetano,A.R., Johnson,R.K. and Pomp,D.
TITLE Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
JOURNAL Unpublished (2001)
COMMENT Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized porcine ovarian follicles library
Seq primer: M13 -29
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .564
/organism="Sus scrofa"

```
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-br-h-08-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_LIB=UNL-P-FN
                     TAG_TISSUE=porcine ovarian follicles
                     TAG_SEQ=CACACT"
BASE COUNT      141 a    129 c    143 g    151 t
ORIGIN


  Query Match             6.4%;  Score 190;  DB 11;  Length 564;
  Best Local Similarity   99.3%;  Pred. No. 5.4e-37;
  Matches  290;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   2467 cctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggcctgcta 2526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    564 CCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTA 505

Qy   2527 tcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggattctgtaa 2586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    504 TCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAA 445

Qy   2587 aactgactggcaccatagaagaattgattacacattgacagccctagtgacctatctcaa 2646
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    444 AACTGACTGGCACCATAGAAGAACTGATTACACATTGACAGCCCTAGTGACCTATCTCAA 385

Qy   2647 ctgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagctattttaa 2706
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    384 CTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGACCTCCAGAGCTATTTTAA 325

Qy   2707 aatgattggtccacggagcgcgggctctcgatattttaaaatgattggtcca 2758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    324 AATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCA 273


RESULT   2
BI360879
LOCUS       BI360879      558 bp    mRNA            EST       01-AUG-2001
DEFINITION  388957 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BI360879
VERSION     BI360879.1  GI:15056907
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 142  row: B  column: 4
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      129 a    123 c    119 g    187 t
ORIGIN


  Query Match             3.1%;  Score 92;  DB 11;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-13;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2845 gcactcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgct 2904
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 GCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGTGTACGACTGTGGGCCCCAGCGCGCT 191

Qy   2905 tggaataaaaatcctcttgctgtttgcatcaa 2936
          ||||||||||||||||||||||||||||||||
Db    192 TGGAATAAAAATCCTCTTGCTGTTTGCATCAA 223


RESULT   3
BF712162/c
LOCUS       BF712162      602 bp    mRNA            EST       02-JAN-2001
DEFINITION  MI-P-E6-acx-e-11-1-UM.s1 MI-P-E6 Sus scrofa cDNA clone
            MI-P-E6-acx-e-11-1-UM 3', mRNA sequence.
ACCESSION   BF712162
VERSION     BF712162.1  GI:12011595
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized fetus at gestational day 20 library cDNA Library
            Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
            Research Center, Department of Animal Science, University of
            Missouri-Columbia, 65211 Clone distribution: clones will be
            available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
```

```
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-E6-acx-e-11-1-UM"
                     /clone_lib="MI-P-E6"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-E6
                     library is derived from fetus at gestational day 20.  For
                     a detailed description of the  library from which this
                     clone was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/.
                     TAG_LIB=MI-P-E6
                     TAG_TISSUE=fetus at gestational day 20
                     TAG_SEQ=AGGAA"
BASE COUNT      141 a    146 c    152 g    162 t      1 others
ORIGIN


  Query Match             3.1%;  Score 92;  DB 11;  Length 602;
  Best Local Similarity   99.3%;  Pred. No. 2.1e-13;
  Matches  142;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    2807 ggctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgcagtcctct 2866
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     143 GGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGTCCTCT 84

Qy    2867 acccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctg 2926
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      83 ACCCCTGCGTGGCGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTG 24

Qy    2927 tttgcatcaaaaaaaaaaaaaaa 2949
           |||||||||||||||||||||||
Db      23 TTTGCATCAAAAAAAAAAAAAAA 1


RESULT   4
AW657531
LOCUS       AW657531      549 bp    mRNA            EST       25-APR-2001
DEFINITION  110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW657531
VERSION     AW657531.1  GI:7423429
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 94  row: E  column: 12
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT      145 a    144 c    121 g    139 t
ORIGIN


  Query Match             2.6%;  Score 78;  DB 10;  Length 549;
  Best Local Similarity   99.2%;  Pred. No. 5.4e-10;
  Matches  128;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1387 aaattcaatgtgaccaaagagcatagaaatcaatgtacatgggggtcccgaaataagctt 1446
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159 AAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTC 218

Qy    1447 accctcactgaagtttccgggaaggggacatgcataggaaaagctccccccatcccaccaa 1506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     219 ACCCTCACTGAAGTTTCCGGGAAGGGGACATGCATAGGAAAAGCTCCCCCCATCCCACCAA 278

Qy    1507 caccttttgc 1515
           ||||||||||
Db     279 CACCTTTGC 287


RESULT   5
BI186066/c
LOCUS       BI186066      468 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-bo-a-05-0-UNL.s2 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-bo-a-05-0-UNL 3', mRNA sequence.
ACCESSION   BI186066
VERSION     BI186066.1  GI:14660475
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail.
            Seq primer: M13 -29
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .468
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-bo-a-05-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
```

polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806 , 1996.
TAG_SEQ=None found"
BASE COUNT 127 a 98 c 97 g 146 t
ORIGIN

Query Match 2.3%; Score 69; DB 11; Length 468;
Best Local Similarity 100.0%; Pred. No. 9.3e-08;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2884 gactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatcaaaaaaaaa 2943
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 69 GACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAA 10

Qy 2944 aaaaaaaaa 2952
|||||||||
Db 9 AAAAAAAAA 1

RESULT 6
BF704354/c
LOCUS BF704354 499 bp mRNA EST 22-DEC-2000
DEFINITION MI-P-O3-aba-g-01-1-UM.s1 MI-P-O3 Sus scrofa cDNA clone MI-P-O3-aba-g-01-1-UM 3', mRNA sequence.
ACCESSION BF704354
VERSION BF704354.1 GI:11989756
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 499)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .499
/organism="Sus scrofa"
/strain="crossbreed"
/db_xref="taxon:9823"
/clone="MI-P-O3-aba-g-01-1-UM"
/clone_lib="MI-P-O3"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-O3 library is derived from ovary at estrus day 12. For a detailed description of the library from which this clone was derived, please visit our web site at http://pigest.genome.iastate.edu/.
TAG_SEQ=None found"
BASE COUNT 132 a 104 c 107 g 152 t 4 others
ORIGIN

Query Match 2.3%; Score 69; DB 11; Length 499;
Best Local Similarity 100.0%; Pred. No. 8.8e-08;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2884 gactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatcaaaaaaaaa 2943
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 69 GACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAA 10

Qy 2944 aaaaaaaaa 2952
|||||||||
Db 9 AAAAAAAAA 1

RESULT 7
BI399852/c
LOCUS BI399852 521 bp mRNA EST 14-AUG-2001
DEFINITION MI-P-AY1-nrv-d-03-0-UI.s1 MI-P-AY1 Sus scrofa cDNA clone MI-P-AY1-nrv-d-03-0-UI 3', mRNA sequence.
ACCESSION BI399852
VERSION BI399852.1 GI:15178913
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 521)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized placenta library cDNA Library Preparation: M.B. Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab, University of Iowa Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .521
/organism="Sus scrofa"
/strain="crossbreed"
/db_xref="taxon:9823"
/clone="MI-P-AY1-nrv-d-03-0-UI"
/clone_lib="MI-P-AY1"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-AY1 library is normalized library derived from the MI-P-AY0 library, ultimately derived from placenta tissue. For a detailed description of the library from which this clone was derived, please visit our web site at http://pigest.genome.iastate.edu/. The procedure used to create this library has been previously described (Bonaldo

```
                    , Lennon and Soares, Genome Research 6: 791-806, 1996)
                    TAG_LIB=MI-P-AY1
                    TAG_TISSUE=placenta
                    TAG_SEQ=ATTGG"
BASE COUNT      142 a    106 c    103 g    170 t
ORIGIN


  Query Match             2.3%;  Score 69;  DB 11;  Length 521;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-08;
  Matches   69;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2884 gactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatcaaaaaaaaa 2943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 GACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAA 10

Qy   2944 aaaaaaaaa 2952
          |||||||||
Db      9 AAAAAAAAA 1


RESULT   8
BI185535
LOCUS       BI185535      730 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-cx-a-05-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-cx-a-05-0-UNL 3', mRNA sequence.
ACCESSION   BI185535
VERSION     BI185535.1  GI:14659944
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 730)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .730
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-cx-a-05-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT      244 a    149 c    161 g    176 t
ORIGIN


  Query Match             2.3%;  Score 69;  DB 11;  Length 730;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-08;
  Matches   69;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2374 aagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaacccaggaagtta 2433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    649 AAGAAGTGGGGAATGAAAGGATGAAAATGCAACCTAACCCTCCCAGAACCCAGGAAGTTA 708

Qy   2434 ataaaaagc 2442
          |||||||||
Db    709 ATAAAAAGC 717


RESULT   9
BI182727/c
LOCUS       BI182727      468 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-bo-a-05-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-bo-a-05-0-UNL 3', mRNA sequence.
ACCESSION   BI182727
VERSION     BI182727.1  GI:14657136
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail.
            Seq primer: M13 -29
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .468
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-bo-a-05-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT      127 a     98 c     97 g    145 t      1 others
ORIGIN


  Query Match             2.3%;  Score 68;  DB 11;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-07;
```

```
  Matches   68;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2884 gactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatcaaaaaaaaa 2943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 GACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAA 9

Qy   2944 aaaaaaaa 2951
          ||||||||
Db      8 AAAAAAAA 1


RESULT  10
BF703671/c
LOCUS       BF703671      485 bp    mRNA            EST       22-DEC-2000
DEFINITION  MI-P-E4-abp-c-04-1-UM.s1 MI-P-E4 Sus scrofa cDNA clone
            MI-P-E4-abp-c-04-1-UM 3', mRNA sequence.
ACCESSION   BF703671
VERSION     BF703671.1  GI:11989079
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized embryo at gestational day 14 library cDNA Library
            Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
            Research Center, Department of Animal Science, University of
            Missouri-Columbia, 65211 Clone distribution: clones will be
            available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .485
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-E4-abp-c-04-1-UM"
                     /clone_lib="MI-P-E4"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-E4
                     library is derived from embryo at gestational day 14.  For
                     a detailed description of the  library from which this
                     clone was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/.
                     TAG_LIB=MI-P-E4
                     TAG_TISSUE=embryo at gestational day 14
                     TAG_SEQ=ACTCAC"
BASE COUNT      142 a    114 c    109 g    119 t      1 others
ORIGIN


  Query Match             2.2%;  Score 64;  DB 11;  Length 485;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-06;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2819 aaccccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcgtgg 2878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 AACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGG 72

Qy   2879 tgta 2882
          ||||
Db     71 TGTA 68


RESULT  11
BF703805/c
LOCUS       BF703805      554 bp    mRNA            EST       22-DEC-2000
DEFINITION  MI-P-O2-abs-a-11-1-UM.s1 MI-P-O2 Sus scrofa cDNA clone
            MI-P-O2-abs-a-11-1-UM 3', mRNA sequence.
ACCESSION   BF703805
VERSION     BF703805.1  GI:11989213
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized ovary at estrus day 5 library cDNA Library
            Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
            Research Center, Department of Animal Science, University of
            Missouri-Columbia, 65211 Clone distribution: clones will be
            available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-O2-abs-a-11-1-UM"
                     /clone_lib="MI-P-O2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-O2
                     library is derived from ovary at estrus day 5.  For a
                     detailed description of the  library from which this clone
                     was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/.
                     TAG_LIB=MI-P-O2
                     TAG_TISSUE=ovary at estrus day 5
                     TAG_SEQ=ATCATC"
BASE COUNT      158 a    128 c    122 g    146 t
ORIGIN


  Query Match             2.2%;  Score 64;  DB 11;  Length 554;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-06;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2819 aaccccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcgtgg 2878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    134 AACCCCATAAAAGCTGTCCCGATTCCGCACTCGGGGCCGCAGTCCTCTACCCCTGCGTGG 75

Qy   2879 tgta 2882
          ||||
Db     74 TGTA 71


RESULT  12
AZ891915/c
LOCUS       AZ891915      451 bp    DNA             GSS       05-MAR-2001
DEFINITION  RPCI-24-172A20.TV RPCI-24 Mus musculus genomic clone RPCI-24-172A20
            , DNA sequence.
ACCESSION   AZ891915
VERSION     AZ891915.1  GI:13210860
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-172A20.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 172  row: A  column: 20
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .451
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-172A20"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      144 a     91 c    106 g    110 t
ORIGIN


  Query Match             1.8%;  Score 52;  DB 13;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 0.0012;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 tgctgctttcccagcctttgttctctaggctcaaggcgctcgagtgggtgag 163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 TGCTGCTTTCCCAGCCTTTGTTCTCTAGGCTCAAGGCGCTCGAGTGGGTGAG 253


RESULT  13
BF701947/c
LOCUS       BF701947      545 bp    mRNA            EST       22-DEC-2000
DEFINITION  MI-P-E3-aag-c-06-1-UM.s1 MI-P-E3 Sus scrofa cDNA clone
            MI-P-E3-aag-c-06-1-UM 3', mRNA sequence.
ACCESSION   BF701947
VERSION     BF701947.1  GI:11987355
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized fetus at gestational day 45 library cDNA Library
            Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
            Research Center, Department of Animal Science, University of
            Missouri-Columbia, 65211 Clone distribution: clones will be
            available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-E3-aag-c-06-1-UM"
                     /clone_lib="MI-P-E3"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-E3
                     library is derived from fetus at gestational day 45.  For
                     a detailed description of the  library from which this
                     clone was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/.
                     TAG_LIB=MI-P-E3
                     TAG_TISSUE=fetus at gestational day 45
                     TAG_SEQ=CGCGAC"
BASE COUNT      138 a    120 c    121 g    166 t
ORIGIN


  Query Match             1.7%;  Score 51;  DB 11;  Length 545;
  Best Local Similarity   100.0%;  Pred. No. 0.0018;
  Matches   51;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2902 gcttggaataaaaatcctcttgctgtttgcatcaaaaaaaaaaaaaaaaaa 2952
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51 GCTTGGAATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAAAAAAAAAA 1


RESULT  14
BI184279/c
LOCUS       BI184279      300 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-ck-g-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-ck-g-02-0-UNL 3', mRNA sequence.
ACCESSION   BI184279
VERSION     BI184279.1  GI:14658688
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail.
            Seq primer: M13 -29
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-ck-g-02-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT       69 a     62 c     71 g     97 t      1 others
ORIGIN


  Query Match             1.7%;  Score 49;  DB 11;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 0.0094;
  Matches   49;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2884 gactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgca 2932
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 GACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCA 18


RESULT  15
BI182742/c
LOCUS       BI182742      494 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-bo-e-05-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-bo-e-05-0-UNL 3', mRNA sequence.
ACCESSION   BI182742
VERSION     BI182742.1  GI:14657151
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized porcine ovarian follicles library
            Seq primer: M13 -29
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-bo-e-05-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_LIB=UNL-P-FN
                     TAG_TISSUE=porcine ovarian follicles
                     TAG_SEQ=CACACT"
BASE COUNT      118 a    117 c    125 g    134 t
ORIGIN


  Query Match             1.6%;  Score 48;  DB 11;  Length 494;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2889 tgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatcaa 2936
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     64 TGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATCAA 17


Search completed: February 24, 2002, 04:28:52
Job time: 24078 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 20:59:59 ; Search time 6550.43 Seconds
(without alignments)
5446.358 Million cell updates/sec

Title: US-09-171-553B-1
Perfect score: 3320
Sequence: 1 gaattcgcggccgcgtcgac.....aaaaaaaaaaaaaaaaaaaa 3320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	385	11.6	576	13	AZ730616	AZ730616 RPCI-24-1
	2	380.8	11.5	864	13	BH056750	BH056750 RPCI-24-9
c	3	373.2	11.2	673	13	AZ329067	AZ329067 1M0053G10
	4	349	10.5	799	11	BI107577	BI107577 602894023
	5	348.8	10.5	705	13	AZ337037	AZ337037 1M0067L03
	6	341.6	10.3	842	13	BH106844	BH106844 RPCI-24-3
	7	340.2	10.2	789	13	AZ715956	AZ715956 RPCI-24-1
	8	338.8	10.2	755	13	AZ715272	AZ715272 RPCI-24-1
	9	336.8	10.1	653	13	BH083773	BH083773 RPCI-24-2
	10	335	10.1	700	13	AZ827524	AZ827524 2M0104E02
c	11	333.2	10.0	689	13	AZ897994	AZ897994 RPCI-24-1
	12	331.4	10.0	817	13	AZ717731	AZ717731 RPCI-24-1
c	13	330.8	10.0	692	13	AZ769626	AZ769626 1M0570J04
c	14	328	9.9	734	13	BH074726	BH074726 RPCI-24-3
	15	327.6	9.9	610	13	AZ432720	AZ432720 1M0218E17
	16	326.6	9.8	685	13	AZ613787	AZ613787 1M0442L18
	17	324.6	9.8	651	13	AZ940771	AZ940771 2M0200G16
c	18	324	9.8	642	13	AZ624191	AZ624191 1M0462K10
c	19	323	9.7	616	13	AZ962118	AZ962118 2M0230N14
	20	322	9.7	656	13	AZ560266	AZ560266 RPCI-23-2
	21	320.6	9.7	652	13	AZ758880	AZ758880 1M0551B09
c	22	319.2	9.6	725	13	AZ866393	AZ866393 2M0176E18
	23	318.4	9.6	668	13	AZ960291	AZ960291 2M0228I18
c	24	316.8	9.5	618	13	AZ466041	AZ466041 1M0276K20
	25	316	9.5	681	13	BH044907	BH044907 RPCI-24-3
	26	314.6	9.5	646	13	BH041228	BH041228 RPCI-24-3
	27	313.8	9.5	639	13	AZ643009	AZ643009 1M0506J17
	28	313.2	9.4	701	13	AZ860071	AZ860071 2M0165H23
	29	313	9.4	702	13	AZ974419	AZ974419 2M0249C06
	30	312.6	9.4	680	13	AZ747906	AZ747906 RPCI-24-6
c	31	310.8	9.4	575	13	AZ479717	AZ479717 1M0300M05
c	32	310.6	9.4	665	13	AZ377711	AZ377711 1M0132L10
c	33	310.4	9.3	656	13	AZ848987	AZ848987 2M0150A17
	34	309.6	9.3	675	13	AZ915310	AZ915310 RPCI-24-1
	35	309.4	9.3	688	13	AZ825369	AZ825369 2M0100I04
	36	308.6	9.3	600	13	AZ987182	AZ987182 2M0269J09
	37	308.4	9.3	691	13	BH034484	BH034484 RPCI-24-3
	38	308.2	9.3	568	13	AZ824214	AZ824214 2M0098N07
	39	308.2	9.3	638	13	AZ435891	AZ435891 1M0223B14
c	40	307.8	9.3	630	13	AZ596966	AZ596966 1M0410K05
	41	306.4	9.2	640	13	AZ863739	AZ863739 2M0173D02
c	42	306.2	9.2	626	11	BG065719	BG065719 H3034D11-
c	43	304.8	9.2	630	13	AZ838774	AZ838774 2M0134D13
c	44	303.4	9.1	586	13	AZ416244	AZ416244 1M0191I20
c	45	302.8	9.1	583	13	AZ415280	AZ415280 1M0190M06

ALIGNMENTS

RESULT 1
AZ730616
LOCUS AZ730616 576 bp DNA GSS 25-JAN-2001
DEFINITION RPCI-24-149D1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-149D1, DNA sequence.
ACCESSION AZ730616
VERSION AZ730616.1 GI:12491859
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 576)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-24-149D1.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 149 row: D column: 1
Seq primer: SP6
Class: BAC ends.
FEATURES Location/Qualifiers

```
     source          1. .576
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-149D1"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      161 a    139 c    121 g    155 t
ORIGIN


  Query Match             11.6%;  Score 385;  DB 13;  Length 576;
  Best Local Similarity   82.5%;  Pred. No. 1.6e-66;
  Matches  453;  Conservative    0;  Mismatches   95;  Indels    1;  Gaps    1;

Qy    2089 ggaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcag 2148
           || ||||| |||||||| |||||    ||||| || || || |||||| ||||||||||
Db       1 GGCATACCCAAGGTAATCGGGTCCATTAATGGACCTGCCTTTGTTGCCAAGGTAAGTCAG 60

Qy    2149 ggactggccaagatattggggattgattggaaactgcattgtgcatacagaccccaaagc 2208
           ||||| |||||||||||||||||||||| |||  | ||||||| |||||| | ||||| |
Db      61 GGACTAGCCAAGATATTGGGGATTGATTAGAAGTTACATTGTGTATACAGGCTCCAAAAC 120

Qy    2209 tcaggacaggtagagaggatgaatagaaccattaaagagacccttaccaaattgaccaca 2268
           |||||||||||||||||||||||||| |||||| |||||||||||||||||||  |  |
Db     121 TCAGGACAGGTAGAGAGGATGAATAAAACCATAAAAGAGACCCTTACCAAATTG-CTGCT 179

Qy    2269 gagactggcattaatgattggatggctctcctgccctttgtgcttttagggtgaggaac 2328
           |||||  || |||||||||||| |||||||| |||||||||| || ||||| || |||
Db     180 ATGACTGCCACTAATGATTGGATAGCTCTCCTACCCTTTGTGCTCTTCAGGGTTAGAAAC 239

Qy    2329 acccctggacagtttgggctgacccccatgaattgctctacgggggaccccccccgttg 2388
           |||||||||||||| | |||| ||||||||||  ||||| |||||  || | || |||
Db     240 ACCCCTGGACAGTTTAGACTGAACCCCTATGAATCACTCTATGGGGGGTCCTCTCCATTG 299

Qy    2389 gcagaaattgcctttgcacatagtgctgatgtgctgctttcccagcctttgttctctagg 2448
           | | |||| |||| || ||||||||||||||||||||||||||||||||||||||||||
Db     300 GTAAAAATAGCCTCTGTACATAGTGCTGATGTGCTGCTTTCCCAGCCTTTGTTCTCTAGT 359

Qy    2449 ctcaaggcgctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactca 2508
           ||||||| |||||| |||||||| |||| ||  | |||||||||||| | | |||||||||
Db     360 CTCAAGGTGCTCGAATGGGTGAGACAGCAAGTATAGAAGCAGCTCCAGAATGCCTACTCA 419

Qy    2509 ggaggagacttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgc 2568
           ||| |||| || |||  ||||||| ||||||||||  ||||||||||||  ||||||||
Db     420 AGAGAAGACCTGGAAGCCCCACATCTCTTCCAAGTTATAGATTCAGTCTACATTAGACGC 479

Qy    2569 caccgtgcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccaca 2628
           | ||   ||| |||| |  |||||| | |||| || |||||||| ||| | || |||||
Db     480 CTCCACACAGAAAACTTTAAGACTCAGCGGAAAGGCCCTTATCTTGTAATCTTTACCACC 539

Qy    2629 ccaacggct 2637
           | ||   ||
Db     540 CTAATTTCT 548


RESULT   2
BH056750
LOCUS       BH056750      864 bp    DNA             GSS       18-JUL-2001
DEFINITION  RPCI-24-94C16.TVB RPCI-24 Mus musculus genomic clone RPCI-24-94C16,
            DNA sequence.
ACCESSION   BH056750
VERSION     BH056750.1  GI:14864050
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 864)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-94C16.TJ RPCI-24-94C16.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 94  row: C  column: 16
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .864
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-94C16"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      237 a    221 c    209 g    197 t
ORIGIN


  Query Match             11.5%;  Score 380.8;  DB 13;  Length 864;
  Best Local Similarity   66.5%;  Pred. No. 1e-65;
  Matches  577;  Conservative    0;  Mismatches  287;  Indels    4;  Gaps    2;

Qy     463 gtgagagagtttttggggacagctggattttgcagactgtggatcccgggggtttgcgacc 522
           ||  | |||||| |||| || || || ||||| ||||| || || || ||||   | |
Db       1 GTACGTGAGTTTCTGGGAACGGCAGGCTTTTGTAGACTCTGAATACCAGGGTCCACAATG 60

Qy     523 ttagcagccccactctacccgctaaccaaagaaaaaggggggattctcctgggctcctgag 582
           ||||| || || || ||||| || || ||||||||||     ||| ||||| |    ||
Db      61 TTAGCGGCACCTCTGTACCCACTCACTAAAGAAAAAGTCCCGTTCACCTGGACAGAAGAA 120

Qy     583 caccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctc 642
           ||||| || |  ||| |    || ||    || |||||    || |  ||| |||||||
Db     121 CACCAAAAAGACTTTAACAGCATAAAGGCTGCTCTGCTTGCAGCCCAGGCTTTGGCCCTG 180

Qy     643 cctgacgtaactaaaccctttaccctttatgtggatgagcgtaagggagtagcccgagga 702
           ||  |  | ||||| || || ||| | || || |||||||     ||||||||||  ||
Db     181 CCATATTTGACTAAGCCTTTCACCATATACGTTGATGAGCCAGCTGGAGTAGCCCGCAGA 240

Qy     703 gttttaacccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagctt 762
           ||  | || || || || || || |||| ||| |  || ||||||||||| || ||  |
Db     241 GTCCTGACTCAGACTCTGGGGCCGTGGAAGAGGCTAGTGGCCTACCTGTCTAAAAAATTG 300

Qy     763 gatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccatactg 822
           || || || |||||||  ||||||   || || || || ||||| || || ||| ||||
Db     301 GACCCAGTTGCCAGTGAATGGCCCTCCTGCCTAAAAGCCATCGCTGCAGTAGCCCTACTT 360

Qy     823 gtcaaggacgctgacaaattgactttgggacagaatataactgtaatagccccccatgca 882
           || || ||  ||| ||| | ||| | || ||| || | | |||||  | |||||||
Db     361 GTTAAAGATACTGGCAAGCTTACTCTAGGTCAGCATGTGAGTGTAATCACTCCCCATGCG 420
```

```
Qy     883 ttggagaacatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacccac 942
           ||  |||  ||  ||  | ||||||||| ||||  |||||||| || ||||| |||||||||
Db     421 TTAAAGAGTATTGTATGACAGCCCCCTGACCATTGGATGACAAATGCCCGGATGACCCAC 480

Qy     943 tatcaaagcctgcttctcacagagagggtcactttcgctccaccagccgctctcaaccct 1002
           || || ||||||  | ||  |  ||  | || |||||  || |  |||  |||    | ||| |||
Db     481 TACCAGAGCCTGTTGCTAAATGACCGAGTGACTTTTGCCCTACCTGCCATCCACAATCCT 540

Qy    1003 gccactctttctgcctgaagagactgatgaaccagtgactcatgattgccatcaactattg 1062
           ||||  |||||  ||  |   |  |||     |   |   ||  | ||   | |  |  ||
Db     541 ACCACCCTTCTCCCCAAGACGGATGACTCTACCTGGGTACACCAATGTGCTGATATCCTG 600

Qy    1063 attgaggagactggggtccgcaaggaccttacagacataccgctgactggagaagtgcta 1122
           |||  ||  |||||     |  |||||  ||  ||  |||  |  |   |  |   ||  ||||
Db     601 GATGAAGAAACTGGATCGAGAAAGGATCTGACTGACCAAACTTGGCC---AGGTGTGCCT 657

Qy    1123 acctggttcactgacggaagcagctatgtggtggaaggtaagaggatggctggggcggca 1182
           |  |||  |||  ||  ||  |||||||  |||||||||  ||  ||  |   |   ||  ||  | ||||
Db     658 GACAGGTACACAGATGGTAGCAGCTTTGTGGTGGAGGGGAAAATAAGAGCAGGAGAGGCA 717

Qy    1183 gtggtggacgggacccgcacgatctgggccagcagcctgccggaaggaaacttcagcgcaa 1242
           |||||  |||  |  |   |      ||  |||||  ||    ||  ||  ||||||||||||  ||  ||
Db     718 GTGGTTGACAGAAAACATGTAATTTGGGCAAGTTAACTTCCAGAAGGAACTTCCGCTCAG 777

Qy    1243 aaggctgagctcatggccctcacgcaagctttgcggctggccgaagggaaatccataaac 1302
           ||||||||||||  ||  |  ||  ||  |||||  |||  |||||  |  ||  ||  |   |  ||  ||
Db     778 AAGGCTGAGCTCCTGACGCTAACTAAAGCTCTGCCGCTGG-CAAAAGGTAGGCAATCAAT 836

Qy    1303 atttatacggacagcaggtatgcctttg 1330
           |||||  ||  ||||||   |||||  ||||
Db     837 ATTTACACTGACAGCCACTATGCTTTTG 864
```

```
RESULT   3
AZ329067/c
LOCUS       AZ329067      673 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0053G10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0053G10 F, DNA sequence.
ACCESSION   AZ329067
VERSION     AZ329067.1  GI:10389414
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0053  row: G  column: 10
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 673.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0053G10"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      165 a    173 c    149 g    186 t
ORIGIN
```

```
  Query Match             11.2%;  Score 373.2;  DB 13;  Length 673;
  Best Local Similarity   78.1%;  Pred. No. 3.5e-64;
  Matches  520;  Conservative    0;  Mismatches  123;  Indels   23;  Gaps    5;

Qy    1900 ctaaggggaagccacccaggcgctcactgggaagtggacttcactgaggtaaagccggct 1959
           ||  ||||   ||   ||  |  ||  ||||||| || || || || ||   |||||  |||
Db     672 CTCAGGGATGGCTGGCCTGAAGCCTACTGGGAGGTAGAGTTTACAGAATCTAAGCCAGCT 613

Qy    1960 aaatacggaaacaaatatctattggttttgtagacacctttcaggatgggtagaggct 2019
           |  ||||||  |||||  ||  ||||  ||  |||  ||||  |||||||||  |  |||  ||||  |||
Db     612 AGATACGGCAACAAGTACTTATTAGTGTTTATAGATACCTTTTCAAGGTGGATAGAAGCT 553

Qy    2020 tatcctactaagaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttt 2079
           |  ||||||  |  ||  |||||              |||  ||  |  ||  ||  ||  |||||  ||
Db     552 TTTCCTACCAGGACTGAGAC------------GGCCAATATGATCCTAGACGAAATCTTC 505

Qy    2080 ccaagatttggaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccag 2139
           ||  ||  ||||  |||||  |||||||  ||||  |  |||||  ||  ||  ||  |||||||||
Db     504 CCGAGGTTTGTGATACCCAAGGTAATCAGGTCCAATAATGGACCTGCCTTTGTTGCCCAG 445

Qy    2140 gtaagtcagggactggccaagatattggggattgattggaaactgcattgtgcatacaga 2199
           ||||||||||||||||||||||||  |  |||||  ||  |||  |  ||||||||||||||
Db     444 GTAAGTCAGGGACTGGCCAAGATATTAGAGATTGGTTAGAAGTTACATTGTGCATACAGG 385

Qy    2200 ccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacccttaccaaa 2259
           ||||||||||||||||||||||||  |  ||||||||||||  ||||||||||||||||
Db     384 CCCCAAAGCTCAGGACAGGTAGAGAGAACGAATAGAACCATT-AAGAGACCCTTACCAAA 326

Qy    2260 ttgaccacagagactggcattaatgattggatggctctcctgccctttgtgctttttagg 2319
           |||  |  |  ||||||  |  |||||||  |||  ||||||||  ||||||||||  ||  |||
Db     325 TTGGCAGCTGAGACTAGAGCTAATGATAAGATAGCTCTCCTACCCTTTGTGCTCTTCAGG 266

Qy    2320 gtgaggaacacccctggacagtttgggctgacccccctatgaattgctctac----ggggg 2375
           ||  ||  ||||  |||||  |||||||||  ||||||||||  ||||  ||||||    |||||
Db     265 GTTAGAAACATCCCTGAACAGTTTGGACTGACCCCCTATAAATTACTCTACGGGGGGGGG 206

Qy    2376 acccccccgttggcagaaattgcctttgcacatagtgctgatgtgctgctttcccagcc 2435
           |  ||  |||  |  ||||  ||||  ||  ||||||||||||  ||||||||||||||||
Db     205 GGGGGCTCCACTGGTAAAAATAGCCTCTGTACATAGTGCTGACATGCTGCTTTCCCAGCC 146

Qy    2436 tttgttctctaggctcaaggcgctcgagtgggtgaggcagcga----gcgtggaagcagc 2491
           ||||||||||||||||||||  ||||||||||||||||  |||  ||    ||  ||||||||||
Db     145 TTTGTTCTCTAGGCTCAAGGTGCTCGAGTGGGTGAGACAGTGAGTGGGCATGGAAGCAGC 86

Qy    2492 tccgggaggcctactcaggag--gagacttgcaagttccacatcgcttccaagttggaga 2549
           |||  |  ||||||||||  |||  |||||  ||||||  ||||||| |||  |||||| ||||
Db      85 TCCAGAAGGCCTACTCAAGAGAAGAGACCTGCAAGCCCCACATCTCTTACAAGTTAGAGA 26
```

```
Qy   2550 ttcagt 2555
          ||| ||
Db     25 TTCGGT 20


RESULT   4
BI107577
LOCUS       BI107577      799 bp    mRNA            EST       26-JUN-2001
DEFINITION  602894023F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5039044 5',
            mRNA sequence.
ACCESSION   BI107577
VERSION     BI107577.1  GI:14558470
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 799)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11107  row: n  column: 05
            High quality sequence stop: 743.
FEATURES             Location/Qualifiers
     source          1. .799
                     /organism="Mus musculus"
                     /strain="CZECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5039044"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      211 a    215 c    218 g    154 t      1 others
ORIGIN


  Query Match           10.5%;  Score 349;  DB 11;  Length 799;
  Best Local Similarity   70.6%;  Pred. No. 2.2e-59;
  Matches  537;  Conservative    0;  Mismatches  215;  Indels    9;  Gaps    5;

Qy    729 gaggagacctgttgcctacctgtcaaagaagcttgatcctgtagccagtggttggccccgt 788
          | || | || || |||||  | |||||||| || ||  |||| |||||||| ||||||
Db     19 GCGGCGCCCGGTAGCCTAGTTATCAAAGAAACTGGACGCTGTGGCCAGTGGATGGCCCTC 78

Qy    789 atgtctgaaggctatcgcagctgtggccatactggtcaaggacgctgacaaattgacttt 848
          || |||   || || |||||   |||  | || || |  || |||||||| ||||| |
Db     79 CTGCCTGCGAGCGATAGCAGCCACGGCTGTGCTAGTAAGAGATGCTGACAAACTGACTAT 138

Qy    849 gggacagaatataactgtaatagccccccatgcattggagaacatcgttcggcagccccc 908
          ||| |||||| | ||| || | ||||| ||  |  | |||| |||| |  |||| || ||
Db    139 GGGCCAGAATGTTACTATAGTGGCCCCACACTCTCTTGAGAGCATCATCAGGCAACCACC 198

Qy    909 agaccgatggatgaccaacgcccgcatgacccactatcaaagcctgcttctcacagagag 968
          ||||| ||||||||||||||||| ||||| ||||| || |||||  | || |||||| |
Db    199 GGACCGCTGGATGACCAACGCCCGAATGACGCACTACCAGAGCCTATTGCTGACAGAGCG 258

Qy    969 ggtcactttcgctccaccagccgctctcaaccctgccactcttctgcctgaagagactga 1028
          || | ||| || || |||||| |||||||| ||| |  | || |||   ||| ||||
Db    259 AGTAAGTTTTGCACCCCCAGCCATTCTCAACCCCGCCTCCTTACTACCT---GAGGCTGA 315

Qy   1029 tgaaccagtgactcatgattgccatcaactattgattgaggagactggggtccgcaagga 1088
          ||  |     | ||| | ||  |  || || ||   || ||||||||  ||||   ||
Db    316 CGAGGCCCCTGCACATAAGTGTGAAGAAATACTGGCAGAAGAGACTGGAATCCGGCCAGA 375

Qy   1089 ccttacagacataccgctgactggagaagtgctaacctggttcactgacggaagcagcta 1148
          ||| ||||||  |||   | | || |   ||   || |||||||| ||||||||||||
Db    376 CCTCACAGACCAACCTTGGCCAGGGGCGATG---ACTTGGTTCACGGACGGAAGCAGCTT 432

Qy   1149 tgtggtggaaggtaagaggatggctggggcggcagtggtggacgggacccgc-acgatct 1207
          |||||| |||||||||| ||| |||||||||| ||||| ||||| || ||  ||    || |
Db    433 TGTGGTAGAAGGTAAGCGGAAGGCTGGGGCTGCAGTAGTGGATGGAACAGGCTGTCATAT 492

Qy   1208 gggccagcagcctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgc 1267
          |||||||||| |||||||| || || ||||| ||||| || || || || || ||  | | |
Db    493 GGGCCAGCAGTCTGCCGGAGGGTACATCAGCTCAAAAAGCGGAACTAATCGCATTAATTC 552

Qy   1268 aagctttgcggctggccgaagggaaatccataaacatttatacggacagcaggtatgcct 1327
          |||| ||  |||||||| ||||| |   |  | ||  | ||||| ||||||  ||| || |
Db    553 AAGCCTTAAGGCTGGCAGAAGGAAGGGCTCTTAATGTCTATACCGACAGCCTGTACGCTT 612

Qy   1328 ttgcgactgcacacgtacacggggccatctataaacaaaggggttgcttacctcagcag 1387
          |||| || || || || ||||| || || ||   |||  | ||  |||| || || |  |
Db    613 TTGCCACGGCTCATGTTCACGGAGCAATATACCGACACCGTGGACTGCTGACGTCTGACG 672

Qy   1388 ggagggaaataaagaaca-aagaggaaattctaagcctattagaagcct-tacatttgcc 1445
          | |  || || || || | | || |||||||| ||| ||||||||||| | ||||||||
Db    673 GCAAAGATATCAAAAATACACGAAGAAATTCTCAGCTTATTAGAAGCCTGTTCATTTGCC 732

Qy   1446 aaaaaggctagctattatacactgtcctggacatcagaaag 1486
             ||| | || ||||| || || || ||||| ||||| |
Db    733 CCGTAGGGTGGCAATTATCCATTGCCCAGGACACCAGAAGG 773


RESULT   5
AZ337037
LOCUS       AZ337037      705 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0067L03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0067L03 R, DNA sequence.
ACCESSION   AZ337037
VERSION     AZ337037.1  GI:10406938
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0067  row: L  column: 03
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 705.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0067L03"
```

/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 184 a 175 c 185 g 161 t
ORIGIN

Query Match 10.5%; Score 348.8; DB 13; Length 705;
Best Local Similarity 69.9%; Pred. No. 2.4e-59;
Matches 484; Conservative 0; Mismatches 207; Indels 1; Gaps 1;

```
Qy    314 taggctacagagcctctgctaagaaggcccagatttgcaggagagaggtaacatacttgg 373
          | || ||| | |||||||||||| || |||||| | ||   ||  || ||    ||| ||
Db      2 TGGGTTACCGGGCCTCTGCTAAAAAAGCCCAGCTATGTCAGATGGAAGTGGTCTACCTGA 61

Qy    374 ggtacagtttgcggggcgggcagcgatggctgacggaggcacggaagaaaactgtagtcc 433
          | || |   | ||||  |   | || |||||||| |||||  | || |    |||    |
Db     62 GATATACCCTCCGGGATGAAAAACGGTGGCTGACAGAGGCTAGAAAAAGGGTTGTGACTC 121

Qy    434 agataccggccccaaccacagccaaacaagtgagagagtttttggggacagctggatttt 493
          | |||||  | ||  | ||  | | ||| |||||||| || |||||  |  |||||  || |
Db    122 AAATACCAACTCCGGCTACTCCAAGACAGGTGAGAGAATTCTTGGGCAACGCTGGGTTCT 181

Qy    494 gcagactgtggatcccggggtttgcgaccttagcagccccactctacccgctaaccaaag 553
          ||||||| ||||| || |||||||| ||  | || ||||| || ||||| || ||||| |
Db    182 GCAGACTTTGGATACCTGGGTTTGCAACACTGGCCGCCCC-CTGTACCCCCTGACCAAGG 240

Qy    554 aaaaagggggattctcctgggctcctgagcaccagaaggcatttgatgctatcaaaaagg 613
          |||  || | |||    ||| |  | || || ||||| || |||||   ||| || ||||
Db    241 AAAGTGGAGAATTTAGATGGACATCAGAACAGCAGAAAGCCTTTGAAAATATTAAGAAGG 300

Qy    614 ccctgctgagcgcacctgctctggccctccctgacgtaactaaaccctttacccttatg 673
          |||||||||| || || ||  |||| || || ||  |||| || ||||| | ||| ||||
Db    301 CCCTGCTGAGTGCGCCAGCCTTGGCACTGCCAGATCTAACCAAGCCCTTCATCCTCTATG 360

Qy    674 tggatgagcgtaagggagtagcccgaggagttttaacccaaaccctaggaccatggagga 733
          | |||||| |  | || || ||| |||||||| | || ||  | ||||| || ||   ||
Db    361 TTGATGAGAGGGAAGGGGTGGCCAGAGGAGTTCTTACTCAGGCTCTAGGGCCGTGAGAGA 420

Qy    734 gacctgttgcctacctgtcaaagaagcttgatcctgtagccagtggttggcccgtatgtc 793
          | || || || ||  |||| |||||||| || || || |||||  | ||||||  |  |
Db    421 GGCCGGTGGCATATTTGTCTAAGAAGCTGGACCCAGTGGCCAGCAGATGGCCCACACATT 480

Qy    794 tgaaggctatcgcagctgtggccatactggtcaaggacgctgacaaattgactttgggac 853
          | |||||  | ||  | || ||| |||| || ||||| |||||||||| ||| ||||||
Db    481 TAAAGGCCGTTGCTACAGTAGCCCTACTCGTTAAGGATGCTGACAAATTAACTCTGGGAC 540

Qy    854 agaatataactgtaatagccccccatgcattggagaacatcgttcggcagcccccagacc 913
          || | ||||||||  ||||||||||  |  | || | ||| |  | || ||||||||||
Db    541 AGCAAATAACTGTGGTAGCCCCCCACTCTCTTGAAAGCATTATCTGTCATCCCCCAGACC 600

Qy    914 gatggatgaccaacgcccgcatgacccactatcaaagcctgcttctcacagagagggtca 973
          |||||||||| || ||| | ||||| ||||| || || ||||| || || ||  |||||
Db    601 GATGGATGACAAATGCCTGGATGACTCACTACCAGAGGCTGCTGCTGACTGAATGGGTCG 660
```

```
Qy    974 ctttcgctccaccagccgctctcaaccctgcc 1005
          || | ||| || ||   ||||||  |||||
Db    661 TGTTTGTTCCTCCTGCTATTCTCAATACTGCC 692
```

RESULT 6
BH106844
LOCUS BH106844 842 bp DNA GSS 19-JUL-2001
DEFINITION RPCI-24-352A17.TVB RPCI-24 Mus musculus genomic clone RPCI-24-352A17, DNA sequence.
ACCESSION BH106844
VERSION BH106844.1 GI:14937281
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 842)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-24-352A17.TJB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 352 row: A column: 17
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .842
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-352A17"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."
BASE COUNT 221 a 228 c 200 g 193 t
ORIGIN

Query Match 10.3%; Score 341.6; DB 13; Length 842;
Best Local Similarity 65.2%; Pred. No. 6.3e-58;
Matches 503; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

```
Qy    332 ctaagaaggcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggcg 391
          |||||| || ||| | |||      ||  | || ||| | || |||    | || |  |
Db     22 CTAAGAAAGCTCAGCTATGCCATTTGGACTTGACCTACCTAGGATACTCCCTCCGAGAAG 81

Qy    392 ggcagcgatggctgacggaggcacggaagaaaactgtagtccagataccggccccaacca 451
          |  |||  ||| | || || || || ||||| |||||  | ||||| ||  |||| ||||
Db     82 GAAAGCAGTGGTTCACCGAAGCTCGAAAGAAGACTGTGATGCAGATCCCAACCCCCACCA 141

Qy    452 cagccaaacaagtgagagagtttttggggacagctggattttgcagactgtggatcccgg 511
          |      ||||||  | |||||| |||| || || || ||||| ||||| || || || |
Db    142 CTCTGTGACAAGTACGTGAGTTTCTGGGAACGGCAGGCTTTTGTAGACTCTGAATACCAG 201
```

```
Qy    512 ggtttgcgaccttagcagccccactctacccgctaaccaaagaaaaaggggattctcct 571
Db    202 GGTCCACAATGTTGGCGGCACCTCTGTACCCACTCACTAAAGAAAAAGTCCCGTTCACCT 261

Qy    572 gggctcctgagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctg 631
Db    262 GGACAGAGGAACACCAAAAAGCCTTTAACAGCATAAAGGCTGCTCTGCTTGCAGCCCAGG 321

Qy    632 ctctggccctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaagggag 691
Db    322 CTTTGGCCCTGCCAGATTTGACTAAGCCTTTCACCATATACGTTGATGAGCCAGCTGGAG 381

Qy    692 tagcccgaggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctgt 751
Db    382 TAGCCCGCAGAGTCCTGACTCAGACTCTGGGGCCCTGGAAGAGGCCAGTGGCCTACCTGT 441

Qy    752 caaagaagcttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctg 811
Db    442 CTAAAAAATTGGACCCAGTTGCCAGTGAATGGCCCTCCTGCCTAAAAGCCATCGCTGTAG 501

Qy    812 tggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaatag 871
Db    502 TAGCCCTCCTTGTTAAAGATACTGGCAAGTTTACTCTAGGTCAGCATGTGAGTTTAATCA 561

Qy    872 ccccccatgcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgccc 931
Db    562 CTCCCCATGCGTTAAAGAGTATTGTATGACAGCCCCCTGACCATTGGATGACAAATGCCC 621

Qy    932 gcatgacccactatcaaagcctgcttctcacagagagggtcactttcgctccaccagccg 991
Db    622 GGATGACCCACTACCAGAGCCTGTTGCTAAATGACCGAGTGACTTTTGCCCTACCTGCCA 681

Qy    992 ctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgcc 1051
Db    682 TCCACAATCCTACCACCCTCCTTCCCAAGACGGATGCCTCTACCTGGGTACACCAATGTG 741

Qy   1052 atcaactattgattgaggagactggggtccgcaaggaccttacagacatacc 1103
Db    742 CTGATATTCTGGCTGAAGAAACTGGATCGAGAAAGGATCTGACTGACCGAAC 793
```

RESULT 7
AZ715956
LOCUS AZ715956 789 bp DNA GSS 24-JAN-2001
DEFINITION RPCI-24-155A1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-155A1, DNA sequence.
ACCESSION AZ715956
VERSION AZ715956.1 GI:12453181
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 789)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-24-155A1.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 155 row: A column: 1
Seq primer: SP6
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .789
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-155A1"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."
BASE COUNT 230 a 200 c 195 g 164 t
ORIGIN

Query Match 10.2%; Score 340.2; DB 13; Length 789;
Best Local Similarity 65.5%; Pred. No. 1.2e-57;
Matches 514; Conservative 0; Mismatches 268; Indels 3; Gaps 1;

```
Qy    789 atgtctgaaggctatcgcagctgtggccatactggtcaaggacgctgacaaattgacttt 848
Db      4 ATGTCTGAAAGCTATTGCAGCAGTAGCCCTGTTGATTAAAGATGCTGACAAATTGACAAT 63

Qy    849 gggacagaatataactgtaatagccccccatgcattggagaacatcgttcggcagccccc 908
Db     64 GGGACAACAGGTGACTGTTGTGGCCCCTCATGCCTTAGAAAGTATCGTGCGTCAGCCACC 123

Qy    909 agaccgatggatgaccaacgcccgcatgacccactatcaaagcctgcttctcacagagag 968
Db    124 TGACAGATGGATGACAAATGCCCGAATGCCACACTATCAGAGCCTGCTGCTAAATGATTG 183

Qy    969 ggtcactttcgctccaccagccgctctcaaccctgccactcttctgcctgaagagactga 1028
Db    184 TGTAACCTTTGCGCCCCCTGCCATCCTCAACCCAGCTACCCTTCTCCCTCTAACAAATGA 243

Qy   1029 tgaaccagtgactcatgattgccatcaactattgattgaggagactggggtccgcaagga 1088
Db    244 TTCCGTCCCAGTACATCAATGTATGGACATCCTCGCTGAAGAAACTGGGACCAGAAGTGA 303

Qy   1089 ccttacagacataccgctgactggagaagtgctaacctggttcactgacggaagcagcta 1148
Db    304 CCTGACTGACCAACCCTGGCCTGG---AGCTCCCAGTTGGTACACGGACGGCAGCAGTTT 360

Qy   1149 tgtggtggaaggtaagaggatggctggggcggcagtggtggacgggacccgcacgatctg 1208
Db    361 CCTGATAGAGGGAAAGCGAAAGGCTGGAGCTGCGGTGGTAGATGGGAAAAAGGTAATTTG 420

Qy   1209 ggccagcagcctgccggaaggaacttcagcgcaaaaggctgagctcatggccctcacgca 1268
Db    421 GGCAAGCGCTTTGCCTGAAGGAACGTCGGCACAAAAGGCTGAACTTATAGCGCTTATTCA 480

Qy   1269 agctttgcggctggccgaagggaaatccataaacatttatacggacagcaggtatgcctt 1328
Db    481 AGCCCTCCGAGAGGCTAAAGGTAAGATCATTAACATCTACACTGACAGCCGCTATGCTTT 540

Qy   1329 tgcgactgcacacgtacacggggccatctataaacaaaggggttgcttacctcagcagg 1388
Db    541 TGCTACAGCACACATCCATGGGGCCATCTACAGGCAGCAAGGGCTATTGACCTCGGCTGG 600

Qy   1389 gagggaaataaagaacaaagaggaaattctaagcctattagaagccttacatttgccaaa 1448
Db    601 TAAAGACATTAAAAACAAAGAAGAAATTCTGGCCCTGTTGGAAGCCATACATGCACCTAA 660

Qy   1449 aaggctagctattatacactgtcctggacatcagaaagccaaagatctcatatctagagg 1508
Db    661 GAAGGTAGCCATCATCCACTGCCCCGGGCACCAAAGAGGAGAAGACTTGGTGGGCCAGGG 720
```

```
Qy   1509 gaaccagatggctgaccgggttgccaagcaggcagcccaggctgttaaccttctgcctat 1568
Db    721 CCAACCGAATGCAGACTCAGTAGCAAAACAAGTTGCTCAAGGGGCCATGATCTTTACTGA 780

Qy   1569 aatag 1573
Db    781 AAAAG 785


RESULT   8
AZ715272
LOCUS       AZ715272      755 bp    DNA             GSS       24-JAN-2001
DEFINITION  RPCI-24-154N5.TJ RPCI-24 Mus musculus genomic clone RPCI-24-154N5,
            DNA sequence.
ACCESSION   AZ715272
VERSION     AZ715272.1  GI:12451823
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 755)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-154N5.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 154  row: N  column: 5
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .755
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-154N5"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      202 a    194 c    190 g    169 t
ORIGIN


  Query Match             10.2%;  Score 338.8;  DB 13;  Length 755;
  Best Local Similarity   68.1%;  Pred. No. 2.3e-57;
  Matches  501;  Conservative    0;  Mismatches  232;  Indels    3;  Gaps    2;

Qy     49 cccactagccaaccacttttttgccttcgaatggagagatccaggtacgggaagaaccggg 108
Db     23 CCCAACAGCCAGCCCTTGTTCGCTTTCGAATGGTGAGACTCCGAGAGTGGACAAGCCGGA 82

Qy    109 cagctcacctggacccgactgccccaagggttcaagaactccccgaccatctttgacgaa 168
Db     83 CAGCTCACATGGACGAGGCTGCCTCAGGGATTCAAGAACTCGCCCACTTTGTTCGATGAA 142

Qy    169 gccctacacagggacctggccaacttcaggatccaacaccctcaggtgaccctcctccag 228
Db    143 GCCCTACACCGAGATCTTGCTCTTTTCCAAGCCAATAACCCACAGGTGATTCTTCTGCAA 202

Qy    229 tacgtggatgacctgcttctggcgggagccaccaaacaggactgcttagaaggtacgaag 288
Db    203 TATGTAGATGACCTGCTCCTAGCTGCAGAAACACGTGAGGACTGTGAAATTGGGACCCAA 262

Qy    289 gcactactgctggaattgtctgacctaggctacagagcctctgctaagaaggcccagatt 348
Db    263 AACCTCCTGGGCGAGTTAGGTAAGCTGGGGTATCGGGCCTCTGCTAAAAAGGCTCAGTTA 322

Qy    349 tgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcgatggctgacg 408
Db    323 TGCCAGATAAAAGTGACCTACCTAGGATATGTCTTGAGAGATGGACAACGGTGGCTCACA 382

Qy    409 gaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaacaagtgaga 468
Db    383 AAAACCAGAAAACAAGCTGTTATGCAGATCCCGACCCCAACCACTGCTCGCCAGGTAAGA 442

Qy    469 gagtttttggggacagctggattttgcagactgtggatcccggggtttgcgaccttagca 528
Db    443 GAGTTCCTGGGGACCGCCGGGTTTTGCAGACTCTGGATTCCCGGATTTGCCACACTAGCA 502

Qy    529 gccccactctacccgctaaccaaagaaaaaggggattctcctgggctcctgagcaccag 588
Db    503 GCTCCCTTGTATCCACTAACCAAAGAGAAAGGGGAATTCACCT-GGATAGAGAACATCAG 561

Qy    589 aaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctccctgac 648
Db    562 CTAGCCTTTGAAACTCTCAAAAAGGCACTGCTGCAGGCTCCGGCATTGGCCCTGCCAGAT 621

Qy    649 gtaactaaaacccttaccctttatgtggatgagcgtaagggagtagcccgaggagtttta 708
Db    622 TTAAACAAACCTTTCACCCTATACATTGATGAAAGAAATGGAGTGGCAAGGGGAGGCCTT 681

Qy    709 acccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagcttgatcct 768
Db    682 ACCCAGGTTTTGGGACCATGGAAGCGCCCGGTAGCCTACTTATAAAGAAA--CTGACGCT 739

Qy    769 gtagccagtggttggc 784
Db    740 GTGGCCAGTGGATGGC 755


RESULT   9
BH083773
LOCUS       BH083773      653 bp    DNA             GSS       18-JUL-2001
DEFINITION  RPCI-24-293G13.TV RPCI-24 Mus musculus genomic clone RPCI-24-293G13
            , DNA sequence.
ACCESSION   BH083773
VERSION     BH083773.1  GI:14903370
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 653)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-293G13.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
```

(pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 293 row: G column: 13
Seq primer: T7
Class: BAC ends.

```
FEATURES             Location/Qualifiers
     source          1. .653
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-293G13"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      167 a    156 c    181 g    149 t
ORIGIN
```

Query Match 10.1%; Score 336.8; DB 13; Length 653;
Best Local Similarity 69.8%; Pred. No. 6e-57;
Matches 455; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

```
Qy    280 ggtacgaaggcactactgctggaattgtctgacctaggctacagagcctctgctaagaag 339
          || ||  ||   || ||    ||| ||  |||| | || ||  |||  || ||||| |||
Db      2 GGAACTGAGAAGCTCCTCGCAGAACTGAGTGACTTGGGGTATCGAGTTTCGGCTAAAAAG 61

Qy    340 gcccagatttgcaggagagaggtaacatacttggggtacagtttgcggggcgggcagcga 399
          || || |||||    |  |||||||| ||| ||||||| |   | || || ||  |  ||
Db     62 GCACAAATTTGTCAAACTGAGGTAACCTACCTGGGGTATACCCTCCGAGGGGGTAAAAGA 121

Qy    400 tggctgacggaggcacggaagaaaactgtagtccagataccggccccaaccacagccaaa 459
          ||||| || ||||| |||||||| |||||  |   ||| ||  | || |||||  |
Db    122 TGGCTCACAGAGGCCCGGAAGAAGACTGTTATGATGATCCCCTCGCCGACCACCCCACGG 181

Qy    460 caagtgagagagtttttggggacagctggattttgcagactgtggatcccggggtttgcg 519
          || ||  | |||||| ||||||| ||||| ||||| ||||| ||||| || || |||||
Db    182 CAGGTACGTGAGTTTCTGGGGACTGCTGGCTTTTGTAGACTCTGGATTCCAGGCTTTGCA 241

Qy    520 accttagcagccccactctacccgctaaccaaagaaaaaggggattctcctgggctcct 579
          | | |||| || || || || ||  | || || |||    |     ||    |||
Db    242 ATCCTAGCGGCACCTCTATATCCTTTGACTAAGGAAGGGGTTCCTTTTGAGTGGAAAGAA 301

Qy    580 gagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggcc 639
          |||||||| |  || ||||| ||||||||   | | ||  |||  || || |  || ||
Db    302 GAGCACCAAAGAGCTTTTGAGGCTATCAAGTCGTCTCTAATGACTGCCCCCGTGCTAGCA 361

Qy    640 ctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaagggagtagcccga 699
          | || ||| | |||||||| |||  ||| |||||||| ||| |   |||  ||||| |
Db    362 TTACCAGACTTGACTAAACCTTTTGTCCTGTATGTGGACGAGAGAGCGGGCATAGCCAGG 421

Qy    700 ggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaag 759
          ||||| || || ||| | || ||||| |||| ||||||||| ||||| ||||||||||
Db    422 GGAGTGTTGACACAAGCACTGGGACCCTGGAAGAGACCTGTAGCCTATCTGTCAAAGAAA 481

Qy    760 cttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccata 819
          | ||||| || || ||||| ||||||  ||| ||||| ||||| || || |||||| |
Db    482 TTAGATCCCGTTGCTAGTGGATGGCCCACATGCCTGAAAGCTATTGCGGCAGTGGCCCTG 541

Qy    820 ctggtcaaggacgctgacaaattgactttgggacagaatataactgtaatagcccccat 879
          ||| |||| || ||||||||||||||  |||||||  |  | ||||   ||||||| |||
Db    542 CTGATCAAAGATGCTGACAAATTGACAATGGGACAACAGGTGACTGGTGTAGCCCCTCAT 601

Qy    880 gcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgccc 931
          || ||||| |  ||||| | |||||| || ||| ||||| |||| || ||||
Db    602 GCCTTGGAAAGTATCGTGCCGCAGCCACCTGACAGATGGTTGACAAATGCCC 653
```

```
RESULT  10
AZ827524
LOCUS       AZ827524      700 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0104E02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0104E02 F, DNA sequence.
ACCESSION   AZ827524
VERSION     AZ827524.1  GI:12997432
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0104  row: E  column: 02
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 700.
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0104E02"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      178 a    182 c    175 g    165 t
ORIGIN
```

Query Match 10.1%; Score 335; DB 13; Length 700;
Best Local Similarity 69.5%; Pred. No. 1.3e-56;
Matches 455; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

```
Qy    460 caagtgagagagtttttggggacagctggattttgcagactgtggatcccggggtttgcg 519
          || ||  | |||||| ||||||| ||||| ||||| ||||| ||||| || || |||||
Db      8 CAGGTACGTGAGTTTCTGGGGACTGCTGGCTTTTGTAGACTCTGGATTCCAGGCTTTGCA 67
```

```
Qy    520 accttagcagccccactctacccgctaaccaaagaaaaaggggattctcctgggctcct 579
Db     68 ACCCTAGCAGCACCTCTATATCCTTTGACTAAGGAAGGGGTTCCTTTCGAGTGGAAAGAA 127

Qy    580 gagcaccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggcc 639
Db    128 GAACACCAGAGAGCTTTTGAGGCTATCAAGTCGTCTCTAATGACTGCCCCCGCGCTAGCA 187

Qy    640 ctccctgacgtaactaaaccctttaccctttatgtggatgagcgtaagggagtagcccga 699
Db    188 TTACCAGACTTGACTAAGCCTTTCGTCCTATATGTGGACGAGAGAGCGGGTGTAGCCAGG 247

Qy    700 ggagttttaacccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaag 759
Db    248 GGAGTGTTGACACAAGCACTGGGACCCTGGAAGAGACCTGTAGCCTATTTGTCAAAAAAA 307

Qy    760 cttgatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccata 819
Db    308 TTAGATCCTGTTGCTAGTGGATGGCCCACATGTCTGAAAGCTATTGCAGCAGTAGCCCTG 367

Qy    820 ctggtcaaggacgctgacaaattgactttgggacagaatataactgtaatagcccccccat 879
Db    368 TTGATCAAAGATGCTGACAAATTGACAATGGGACAACAGGTGACTGTTGTGGCCCCTCAT 427

Qy    880 gcattggagaacatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacc 939
Db    428 GCCTTAGAAAGTATCGTGCGGCAGCCACCTGACAGATGGATGACAAATGCCCGAATGACA 487

Qy    940 cactatcaaagcctgcttctcacagagagggtcactttcgctccaccagccgctctcaac 999
Db    488 CACTATCAGAGCCTGCTGCTAAATGAGCGTGTAACCTTTGCGCCCCCTGCCATCCTCAAC 547

Qy   1000 cctgccactcttctgcctgaagagactgatgaaccagtgactcatgattgccatcaacta 1059
Db    548 CCAGCTACCCTTCTCCCTCTAACAAATGATTCCGTCCCAGTACATCAATGTATGGACATC 607

Qy   1060 ttgattgaggagactggggtccgcaaggaccttacagacataccgctgactggag 1114
Db    608 CTTGCTGAAGAAACTGGGACCAGAAGTGACCTGACTGACCAACCCTGGCCTGGAG 662
```

```
RESULT  11
AZ897994/c
LOCUS       AZ897994      689 bp    DNA             GSS       05-MAR-2001
DEFINITION  RPCI-24-173A19.TV RPCI-24 Mus musculus genomic clone RPCI-24-173A19
            , DNA sequence.
ACCESSION   AZ897994
VERSION     AZ897994.1  GI:13216939
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-173A19.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 173  row: A  column: 19
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-173A19"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      168 a    174 c    170 g    177 t
ORIGIN
```

```
  Query Match           10.0%;  Score 333.2;  DB 13;  Length 689;
  Best Local Similarity   68.9%;  Pred. No. 3.1e-56;
  Matches  474;  Conservative    0;  Mismatches  208;  Indels    6;  Gaps    1;

Qy   1933 gtggacttcactgaggtaaagccggctaaatacggaaacaaatatctattggtttttgta 1992
Db    688 GTAGATTTTACTGAAGTTAAACCTGGAATGTATGGTAACAAGTATCTGTTAGTATTTGTA 629

Qy   1993 gacacctttttcaggatgggtagaggcttatcctactaagaaagagacttcaaccgtggtg 2052
Db    628 GACACCTTTTTAGGATGGGTCGAGGCATTTCCCACTAAAACAGAGACTGCCCAGATTGTG 569

Qy   2053 gctaagaaaatactggaggaaatttttccaagatttggaatacctaaggtaatagggtca 2112
Db    568 GCCAAGAAGATCCTTGAAGAAATCCTGCCAAGATTTGGAATCCCTAAGGTAATCGGGTCC 509

Qy   2113 gacaatggtccagctttcgttgcccaggtaagtcagggactggccaagatattggggatt 2172
Db    508 GACAATGGACCAGCCTTTGTTGCCCAGGTAAGTCAGGGCTTGGCCACTCAGTTGGGCATC 449

Qy   2173 gattggaaaactgcattgtgcatacagaccccaaagctcaggacaggtagagaggatgaat 2232
Db    448 GATTGGAAATTACACTGTGCTTACCGCCCTCAAAGCTCAGGACAGGTAGAGAGGATGAAT 389

Qy   2233 agaaccattaaagagaccctaccaaattgaccacagagactggcattaatgattggatg 2292
Db    388 AGGACCTTAAAAGAGACCTTGACTAAATTAGCCATTGAGACCGGCGGGAAAGACTGGGTG 329

Qy   2293 gctctcctgccctttgtgcttttagggtgaggaacacccctggacagtttgggctgacc 2352
Db    328 GCTCTCCTCCCTCTTGCGCTCTTCCGAGCCCGGAACACTCCTGGACGTTTCGGGCTCACT 269

Qy   2353 ccctatgaattgctctacgggggaccccccccgttggcagaaattg------cctttgca 2406
Db    268 CCTTTTGAAGTTCTGTATGGAGGACCTCCCCCCTTAATGGAAGCTGGTAGAACATTAGTT 209

Qy   2407 catagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcgagtgg 2466
Db    208 TCCGACTCTGACCCTGTCTTACCCTCCTCTTTGCTTATTCAAATAAAGGCCCTAGAAGTG 149

Qy   2467 gtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagacttgcaagtt 2526
Db    148 ATTAGAACCCAGATTTGGGACCAACTGAAAGCAGCCTATACCCCAGGGACCACCGCAGTA 89

Qy   2527 ccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaaacctc 2586
Db     88 CCCCACGGGTTCCGAGTTGGAGACAAAGTCTTGGTCAGGCGGCATAGAACCGGCAGCCTT 29

Qy   2587 gagactcggtggaagggaccttatctcg 2614
Db     28 GAGCCACGGTGGAAGGGACCCTATTTGG 1
```

```
RESULT  12
AZ717731
LOCUS       AZ717731      817 bp    DNA             GSS       24-JAN-2001
DEFINITION  RPCI-24-155I1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-155I1,
            DNA sequence.
ACCESSION   AZ717731
VERSION     AZ717731.1  GI:12456936
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 817)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-155I1.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 155  row: I  column: 1
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .817
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-155I1"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      243 a    209 c    204 g    161 t
ORIGIN
```

```
  Query Match           10.0%;  Score 331.4;  DB 13;  Length 817;
  Best Local Similarity   65.5%;  Pred. No. 6.7e-56;
  Matches  501;  Conservative    0;  Mismatches  261;  Indels    3;  Gaps    1;

Qy    809 ctgtggccatactggtcaaggacgctgacaaattgactttgggacagaatataactgtaa 868
          | || ||| |  || | || || ||||||||||||||  |||||||  |  | |||||
Db      1 CAGTAGCCCTGTTGATTAAAGATGCTGACAAATTGACAATGGGACAACAGGTGACTGTTG 60

Qy    869 tagccccccatgcattggagaacatcgttcggcagcccccagaccgatggatgaccaacg 928
          | ||||| ||||| || || |  ||||  || ||||| || ||| ||||||||||| || |
Db     61 TGGCCCCTCATGCCTTAGAAAGTATCGCGCGTCAGCCACCTGACAGATGGATGACAAATG 120

Qy    929 cccgcatgacccactatcaaagcctgcttctcacagagagggtcactttcgctccaccag 988
          |||| ||| | |||||||| |||||||| || |  ||| | || || || || || | |
Db    121 CCCGAATGCCACACTATCAGAGCCTGCTGCTAAATGAGCGTGTAACCTTTGCGCCCCCTG 180

Qy    989 ccgctctcaaccctgccactcttctgcctgaagagactgatgaaccagtgactcatgatt 1048
          ||   |||||||| || || ||||| |||  |   | ||||          ||| | |
Db    181 CCATCCTCAACCCAGCTACCCTTCTCCCTCTAACAAATGATTCCGTCCCAGTACATCAAT 240

Qy   1049 gccatcaactattgattgaggagactggggtccgcaaggaccttacagacataccgctga 1108
```

```
          |     |  |  |   ||| || ||||||  | | |  ||||| || |||  |||   |
Db    241 GTATGGACATCCTCGCTGAAGAAACTGGGACCAGAAGTGACCTGACTGACCAACCCTGGC 300

Qy   1109 ctggagaagtgctaacctggttcactgacggaagcagctatgtggtggaaggtaagagga 1168
          ||||   ||  |  |  |||| ||| ||||| ||||| |   || | || || ||| | |
Db    301 CTGG---AGCTCCCAGTTGGTACACGGACGGCAGCAGTTTCCTGATAGAGGGAAAGCGAA 357

Qy   1169 tggctggggcggcagtggtggacgggacccgcacgatctgggccagcagcctgccggaag 1228
          ||||||  ||  || |||||  || ||||          ||  ||||| |||    |||| ||||
Db    358 AGGCTGGAGCTGCGGTGGTAGATGGGAAAAAGGTAATTTGGGCAAGCGCTTTGCCTGAAG 417

Qy   1229 gaacttcagcgcaaaaggctgagctcatggccctcacgcaagctttgcggctggccgaag 1288
          ||||  ||  ||  |||||||||||  ||  ||  ||  ||  |  |||||  |  ||   |||  |||
Db    418 GAACGTCGGCACAAAAGGCTGAACTTATAGCGCTTATTCAAGCCCTCCGAGAGGCTAAAG 477

Qy   1289 ggaaatccataaacatttatacggacagcaggtatgcctttgcgactgcacacgtacacg 1348
          | ||   |||  |||||  ||  ||  ||||||  |  |||||  |||||  ||  ||||||  | ||  |
Db    478 GTAAGATCATTAACATCTACACTGACAGCCGCTATGCTTTTGCTACAGCACACATCCATG 537

Qy   1349 gggccatctataaacaaaggggttgcttacctcagcagggagggaaataaagaacaaag 1408
          |||||||||  |  ||    |||  |  | |||||  ||  ||  |  ||  ||  ||  |||||||
Db    538 GGGCCATCTACAGGCAGCAAGGGCTATTGACCTCGGCTGGTAAAGACATTAAAAACAAAG 597

Qy   1409 aggaaattctaagcctattagaagccttacatttgccaaaaaggctagctattatacact 1468
          |  ||||||||   |||  ||  ||||||  |||||   ||  ||  |  |  ||||  ||  ||  ||||
Db    598 AAGAAATTCTGGCCCTGTTGGAAGCCATACATGCACCTAAGAAGGTAGCCATCATCCACT 657

Qy   1469 gtcctggacatcagaaagccaaagatctcatatctagagggaaccagatggctgaccggg 1528
          | ||  ||  ||  ||  |  ||   ||||  |  |  |  ||  ||  ||||  |||||  |||   |
Db    658 GCCCCGGCCACCAAAGAGGAGAAGACTTGGTGGCCAGGGGCAACCGAATGGCAGACTCAG 717

Qy   1529 ttgccaagcaggcagcccaggctgttaaccttctgcctataatag 1573
          |  ||  ||  ||    ||  ||  |  |  |   |  |  ||  ||  ||
Db    718 TAGCAAAACAAAGTGCTCAAGGGGCCATGATCTTAACTGAAAAAG 762
```

```
RESULT  13
AZ769626/c
LOCUS       AZ769626      692 bp    DNA             GSS       16-FEB-2001
DEFINITION  1M0570J04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0570J04 R, DNA sequence.
ACCESSION   AZ769626
VERSION     AZ769626.1  GI:12889953
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0570  row: J  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 692.
FEATURES             Location/Qualifiers
     source          1. .692
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
```

/db_xref="taxon:10090"
/clone="UUGC1M0570J04"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 171 a 165 c 164 g 192 t
ORIGIN

Query Match 10.0%; Score 330.8; DB 13; Length 692;
Best Local Similarity 69.1%; Pred. No. 9.1e-56;
Matches 452; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

```
Qy   1745 taacccacctaggaactaaacacctgcagcagttggtcagaacatccccttatcatgttc 1804
          | ||||| || |||  | || |  ||  |    |  | | ||  |||  || || || | |
Db    691 TCACCCATCTGGGAGTTGAAAAAATGATGAGACTAATTAAAATTCCCCGATACCAAGTCC 632

Qy   1805 tgaggctaccaggagtggctgactcggtggtcaaacattgtgtgccctgccagctggtta 1864
            |  ||      |||||||  |   | |  |  |    ||    || ||     |   ||
Db    631 CCAACCTGAAGTCAGTGGCTCAAAAGATTATAGACTCCTGCAAACCATGTGCATTCACTA 572

Qy   1865 atgctaatccttccagaatacctccaggaaagagactaaggggaagccacccaggcgctc 1924
          |||| | |    ||   | |   || |||||||||| | |||||  ||  || || | |
Db    571 ATGCGACTAGAGCCTACAAAGAACCTGGAAAGAGACAACGGGGAGACCGTCCTGGAGTGT 512

Qy   1925 actgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatctattgg 1984
          | ||||| || || || ||||| || || || |  |  || || ||||| ||||| || | |
Db    511 ATTGGGAGGTAGATTTTACTGAAGTTAAACCTGGAATGTATGGTAACAAGTATCTGTTAG 452

Qy   1985 tttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagacttcaa 2044
          | |||||||||||||||||||||||||| ||||| | ||| ||||| |  |||||| |
Db    451 TATTTGTAGACACCTTTTCAGGATGGGTTGAGGCGTTTCCCACTAAAACTGAGACTGCCC 392

Qy   2045 ccgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaaggtaa 2104
             | ||||| ||||| || || || |||||  | ||||||||||||||| ||||||||||
Db    391 AGATTGTGGCCAAGAAGATCCTTGAAGAAATCCTGCCAAGATTTGGAATCCCTAAGGTAA 332

Qy   2105 tagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaagatat 2164
          | ||||| |||||||| ||||| || ||||||||||||||||||  ||||||      |
Db    331 TCGGGTCCGACAATGGACCAGCCTTTGTTGCCCAGGTAAGTCAGGGCTTGGCCACTCAGT 272

Qy   2165 tggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggtagaga 2224
          |||| || ||||||||| | || ||||| |||   || ||||||||||||||||||||||
Db    271 TGGGCATCGATTGGAAATTACACTGTGCTTACCACCCTCAAAGCTCAGGACAGGTAGAGA 212

Qy   2225 ggatgaatagaaccattaaagagacccttaccaaattgaccacagagactggcattaatg 2284
          |||||||||| ||| | |||||||| | || |||||  |||  |||||  ||   || | |
Db    211 GGATGAATAGGACCTTAAAAGAGACCTTGACTAAATTAGCCATTGAGACCTGCGGGAAAG 152

Qy   2285 attggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacagtttg 2344
          | ||| |||||||||| ||  ||| ||| ||  | |   | ||||||||||||||  ||
Db    151 ACTGGGTGGCTCTCCTCCCTCTTGCGCTCTTCCGAGCCCGAAACACCCCTGGACATTTCA 92

Qy   2345 ggctgacccccctatgaattgctctacgggggaccccccccgttggcagaaattg 2398
```

```
          |||| || || | |||| | || || || ||||| ||||| ||    ||| ||
Db     91 GGCTCACTCCTTTTGAAGTTCTGTATGGAGGACCTCCCCCCTTAATGGAAGCTG 38
```

RESULT 14
BH074726/c
LOCUS BH074726 734 bp DNA GSS 18-JUL-2001
DEFINITION RPCI-24-331M11.TVB RPCI-24 Mus musculus genomic clone RPCI-24-331M11, DNA sequence.
ACCESSION BH074726
VERSION BH074726.1 GI:14894323
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 734)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-24-331M11.TJB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 331 row: M column: 11
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .734
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-331M11"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."
BASE COUNT 182 a 184 c 182 g 186 t
ORIGIN

Query Match 9.9%; Score 328; DB 13; Length 734;
Best Local Similarity 68.9%; Pred. No. 3.2e-55;
Matches 481; Conservative 0; Mismatches 210; Indels 7; Gaps 2;

```
Qy   1925 actgggaagtggacttcactgaggtaaagccggctaaatacggaaacaaatatct-attg 1983
          | ||| | || || ||  || | || || || |  |  || || | ||| ||||| | |
Db    728 ATTGGAAGGTAGATTTTCCTAAAGTTAAACCTGGAATGTATGGTACCAAGTATCTGAATA 669

Qy   1984 gtttttgtagacacctttttcaggatgggtagaggcttatcctactaagaaagagacttca 2043
          || |||||||||| |||||||||||||| ||||| | ||| ||||| |  |||||| |
Db    668 GTATTTGTAGACACTTTTTCAGGATGGGTTGAGGCGTTTCCCACTAAAACTGAGACTGCC 609

Qy   2044 accgtggtggctaagaaaatactggaggaaatttttccaagatttggaatacctaaggta 2103
             | ||||| ||||| || || || |||||  | |||||||||||||| |||||||||
Db    608 CAGATTGTGGCCAAGAAGATCCTTGAAGAAATCCTGCCAAGATTTGGAATCCCTAAGGTA 549

Qy   2104 atagggtcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaagata 2163
```

```
Db     548 ATCGGGTCCGACAATGGACCAGCCTTTGTTGCCCAGGTAAGTCAGGGCTTGGCCACTCAG 489

Qy    2164 ttggggattgattggaaactgcattgtgcatacagaccccaaagctcaggacaggtagag 2223
Db     488 TTGGGCATCGATTGGAAATTACACTGTGCTTACCGCCCTCAAAGCTCAGGACAGGTAGAG 429

Qy    2224 aggatgaatagaaccattaaagagacccttaccaaattgaccacagagactggcattaat 2283
Db     428 AGGATGAATAGGACCTTAAAAGAGACCTTGACTAAATTAGCCATTGAGACCGGCGGGAAA 369

Qy    2284 gattggatggctctcctgccctttgtgctttttagggtgaggaacacccctggacagttt 2343
Db     368 GACTGGGTGGCTCTCCTCCCTCTAGCGCTCTTCCGAGCCCGTAACACCCCTGGACGTTTC 309

Qy    2344 gggctgacccccctatgaattgctctacgggggaccccccccgttggcagaaattg----- 2398
Db     308 GGGCTCACTCCTTTTGAAGTTCTGTATGGAGGACCTCCCCCCTTAATGGAAGCTGGTGGA 249

Qy    2399 -cctttgcacatagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcg 2457
Db     248 ACATTGGTTTCCGACTCTGACCCTGTCTTACCCTCCTCTTTGCTTATTCATTTAAAGGCC 189

Qy    2458 ctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactcaggaggagac 2517
Db     188 CTAGAAGTGATTAGGACCCAAATTTGGGACCAACTGAAGGCAGCCTATACCCCAGGGACC 129

Qy    2518 ttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgca 2577
Db     128 ACCGCAGTACCCCACGGGTTCCGAGTTGGAGACAAAGTCTTGGTCAGACGGCATCGAACC 69

Qy    2578 ggaaacctcgagactcggtggaagggaccttatctcgt 2615
Db      68 GGCAGCCTCGAGCCACGGTGGAAGGGACCCTATTTGGT 31
```

```
RESULT  15
AZ432720
LOCUS       AZ432720      610 bp    DNA              GSS       03-OCT-2000
DEFINITION  1M0218E17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0218E17 F, DNA sequence.
ACCESSION   AZ432720
VERSION     AZ432720.1  GI:10556733
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0218  row: E  column: 17
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 610.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0218E17"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      170 a    148 c    133 g    159 t
ORIGIN
```

Query Match 9.9%; Score 327.6; DB 13; Length 610;
Best Local Similarity 82.6%; Pred. No. 4.1e-55;
Matches 436; Conservative 0; Mismatches 84; Indels 8; Gaps 5;

```
Qy    2110 tcagacaatggtccagctttcgttgcccaggtaagtcagggactggccaagatattgggg 2169
Db       1 TCCAATAACAGACCTGCATTTGTTGCCCAGGTAAATCAGGGACTGGCCAAGATATTGGGG 60

Qy    2170 attgattggaaactgcattgtgcatacagaccccaaagctcaggacaggtagagaggatg 2229
Db      61 ATTGATTAGAAGTTACATTGTGCATACAGGCCCCAAAGCTCAGGACAGGTAGAGAGAATG 120

Qy    2230 aatagaaccattaaagagacccttaccaaattgaccacagagactggcattaatgattgg 2289
Db     121 AACAGAACCATT-AAGAGACCCTTACCAAATTGACAGCAGAGACTGGCGCTAATGATTGG 179

Qy    2290 atggctctcctgccctttgtgcttttagggtgaggaacacccctggacagtttgggctg 2349
Db     180 ATAGCTCTCCTACCCTTTGTGCTCTTCAGGGTTAAAAACACCCCTGGACAGTTTGGACTG 239

Qy    2350 acccccctatgaattgctctac--gggggacccccccgttggcagaaattgcctttgcac 2407
Db     240 A-CCCCTATAAATCACTCTACGGGGGGGGGGAGCTCCACTGGTAAAAATAGCCTCTGTAC 298

Qy    2408 atagtgctgatgtgctgctttcccagcctttgttctctaggctcaaggcgctcgagtggg 2467
Db     299 ATAGTGCTGACATGCTGCTTTCCACGCCTTTCTTCTCTAGGCTCGAGGTGCTCAAGTGGG 358

Qy    2468 tgaggcagcgagcgtggaagcagctccgggaggcctactcaggag---gagacttgcaag 2524
Db     359 TGAGACAGCAAGCGAAGAAGTAGCTCCAGAAGGCCTACTCAAGAGAAAAAGACCTGCAAG 418

Qy    2525 ttccacatcgcttccaagttggagattcagtctatgttagacgccaccgtgcaggaaacc 2584
Db     419 CCCCACATCGCTTCCAAGTTAGAGATTCGTTCTACATTAGACGTCTCCACGCAGAAAACC 478

Qy    2585 tcgagactc-ggtggaagggaccttatctcgtacttttgaccacacca 2631
Db     479 TTGAGACTCTGTTGGAAAGGCCCTTATCTTGTACTCTTAACCACCCTA 526
```

Search completed: February 23, 2002, 23:36:01
Job time: 9362 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:36:01 ; Search time 6550.43 Seconds
(without alignments)
13466.612 Million cell updates/sec

Title: US-09-171-553B-3
Perfect score: 8209
Sequence: 1 gtggtgtacgactgtgggcc.....aaaaaaaaaaaaaaaaaaaa 8209

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	600	7.3	2085	12	AK014027	AK014027 Mus muscu
	2	594.6	7.2	857	11	BI184013	BI184013 UNL-P-FN-
	3	552.4	6.7	2096	12	AK011989	AK011989 Mus muscu
	4	473.6	5.8	549	10	AW657531	AW657531 110922 MA
	5	472	5.7	2935	12	AK004927	AK004927 Mus muscu
	6	420.6	5.1	467	11	BF441465	BF441465 257856 MA
	7	419.8	5.1	466	11	BF441468	BF441468 257862 MA
c	8	419.4	5.1	465	11	BI399234	BI399234 MI-P-AY1-
	9	404.2	4.9	801	13	BH063533	BH063533 RPCI-24-3
	10	401.6	4.9	452	11	BF441466	BF441466 257857 MA
	11	385	4.7	576	13	AZ730616	AZ730616 RPCI-24-1
c	12	383.6	4.7	564	11	BI182930	BI182930 UNL-P-FN-
	13	380.8	4.6	864	13	BH056750	BH056750 RPCI-24-9
	14	380.4	4.6	729	13	AZ318021	AZ318021 1M0036O04
c	15	373.2	4.5	673	13	AZ329067	AZ329067 1M0053G10
c	16	371	4.5	447	11	BI401063	BI401063 MI-P-CP0-
	17	367	4.5	679	13	BH121259	BH121259 RPCI-24-3
	18	366.6	4.5	646	13	AZ967275	AZ967275 2M0238G10
c	19	358.6	4.4	602	11	BF712162	BF712162 MI-P-E6-a
	20	355.4	4.3	564	11	BI181099	BI181099 UNL-P-FN-
c	21	355.4	4.3	685	13	AZ636636	AZ636636 1M0495O02
	22	354.8	4.3	564	11	BI186129	BI186129 UNL-P-FN-
	23	352.4	4.3	644	13	BH124013	BH124013 RPCI-24-2
	24	352	4.3	376	11	BF441469	BF441469 257863 MA
	25	351.4	4.3	690	13	AZ627566	AZ627566 1M0469L20
	26	349	4.3	799	11	BI107577	BI107577 602894023
	27	348.8	4.2	705	13	AZ337037	AZ337037 1M0067L03
	28	343.8	4.2	598	13	AZ959941	AZ959941 2M0227K18
	29	343.2	4.2	791	13	BH090616	BH090616 RPCI-24-3
	30	341.6	4.2	842	13	BH106844	BH106844 RPCI-24-3
	31	340.2	4.1	789	13	AZ715956	AZ715956 RPCI-24-1
	32	340.2	4.1	975	11	BF784826	BF784826 602111174
	33	338.8	4.1	755	13	AZ715272	AZ715272 RPCI-24-1
	34	338.6	4.1	601	13	AZ650638	AZ650638 1M0520L20
c	35	338	4.1	617	13	AZ420279	AZ420279 1M0198K05
	36	336.8	4.1	653	13	BH083773	BH083773 RPCI-24-2
	37	336.6	4.1	632	10	AL134888	AL134888 DKFZp762K
	38	336.6	4.1	657	13	BH096361	BH096361 RPCI-24-3
	39	336.4	4.1	813	11	BI152889	BI152889 602918048
c	40	336.2	4.1	682	13	AZ980516	AZ980516 2M0257O06
	41	335	4.1	700	13	AZ827524	AZ827524 2M0104E02
c	42	333.2	4.1	689	13	AZ897994	AZ897994 RPCI-24-1
	43	331.4	4.0	817	13	AZ717731	AZ717731 RPCI-24-1
c	44	330.8	4.0	692	13	AZ769626	AZ769626 1M0570J04
c	45	330	4.0	688	13	AZ912534	AZ912534 RPCI-24-1

ALIGNMENTS

RESULT 1
AK014027
LOCUS AK014027 2085 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus 13 days embryo head cDNA, RIKEN full-length enriched library, clone:3110009E22, full insert sequence.
ACCESSION AK014027
VERSION AK014027.1 GI:12851651
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 13 days embryo head cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:3110009E22.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2085)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 2085)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 2085)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (bases 1 to 2085)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 2085)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 50.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES Location/Qualifiers
source 1. .2085
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1906029"
/db_xref="MGD:MGI:1920342"
/clone="3110009E22"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
CDS 356. .1147
/note="putative"
/codon_start=1
/protein_id="BAB29122.1"
/db_xref="GI:12851652"
/translation="MGQTVTTPLSLTLDHWSEIRSRAHNLSVEIKKGPWRTFCASEWP TFDVGWPPEGTFDLTVIFEVKAIVFQDGPGSHPDQQPYITVWQDLVQNSPPWIKPWIP KRKPGSRALALRRPKVKPKDPPEKEPLPGPSALPKIYPEIEGPPEWPEPPPPYPLPPP ASTPPQLPEPGPGAGGGTVGGPAAGTQSRSAQSLSADGVPDTMSAFPLCAVGPAADLH SLQTLQYWPFSSADLYNWKSYFFRKSSWTNQPGGVINVISSAHLG"
BASE COUNT 544 a 504 c 573 g 464 t
ORIGIN

Query Match 7.3%; Score 600; DB 12; Length 2085;

Best Local Similarity 60.6%; Pred. No. 2.4e-111;
Matches 1265; Conservative 0; Mismatches 740; Indels 81; Gaps 14;

```
Qy   251 ttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcctctcagaccgt 310
             |||  | | || ||||  ||||          ||  ||  |  | |  | |           ||
Db    31 TTCAAACTTTCGGTTTGCGGTCGCAACATCTCACGTCTCACGGGAGACAGAGACCCGAGT 90

Qy   311 aaggactggaggactgtgatcagcagacgtgctaggaggatcacaggctgccaccctggg 370
           ||      ||| ||||||| ||||| |||||   |||||||| | ||||| | |||||
Db    91 CTCGAGACATGGAGTGTGATCGGCAGATGTGCT---AGGATCACTGACTGCCGCTCTGGG 147

Qy   371 ggacgccccgggaggtggggagagccagggacgcctggtggtctcctactgtcggtcaga 430
           ||| ||| |||     |||||  | |          |   | ||   || | |   |||
Db   148 GGATGCCTCGGAGCAAGGGGAAGGACTCTAGG---AGTCCGACTGGGACGGGAGTAAAGA 204

Qy   431 ggaccgagttctgttgttgaagcgaaagcttcccccctccgcggccgtccgactcttttgc 490
             |||  |  ||     |  ||     | ||  | || || || |  ||||| |
Db   205 CTCGGTAGTCTTCCTGGATTGGAAGAACTACTCACCAGCCCGACCTTCTGCATCTTTGGA 264

Qy   491 ctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggtttctgtctcgtg 550
           || || ||   | |   || | | |     | | |   ||| |  || | | |||  ||
Db   265 TTGGTTTTGTCTGCCTAAGAGGTGGCAGACATTTGTTAGTCTTTGTGTGTGTTTCTTCTG 324

Qy   551 tgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacagtgactacccc 610
           ||||| | |  |||       || |   ||  | |  |||||||||| ||||| |||||
Db   325 TGTCTGTTTTGTGT------TTTTTGTCACCTGGACGATGGGACAGACTGTGACGACCCC 378

Qy   611 ccttagtttgactctcgaccattggactgaagttagatccagggctcataatttgtcagt 670
           ||| ||||||||  | || || ||| ||||| |||| || ||||| || ||| | |||||
Db   379 CCTAAGTTTGACCTTAGATCACTGGTCTGAAATTAGGTCGAGGGCCCACAATCTTTCAGT 438

Qy   671 tcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaacattcgatgt 730
             || | ||||||||||||||||| |||||| ||||| ||||| |||||||||| ||||
Db   439 GGAGATCAAGAAGGGACCTTGGCGGACTTTTTGTGCTTCTGAGTGGCCAACATTTGATGT 498

Qy   731 tggatggccatcagaggggaccttttaattctgaaattatcctggctgttaaggcaatcat 790
           || |||||| |||||||||| ||| |          | ||| | |  ||||| || ||  |
Db   499 GGGCTGGCCACCAGAGGGGACTTTTGACCTGACTGTGATCTTTGAAGTTAAAGCCATTGT 558

Qy   791 ttttcagactggacccggctctcatcctgatcaggagccctatatccttacgtggcaaga 850
           ||||||  |||||| || || || || || ||| | ||||| ||   |  |||||||||
Db   559 CTTTCAGGATGGACCTGGGTCCCACCCAGACCAGCAACCCTACATAACTGTGTGGCAAGA 618

Qy   851 tttggcagaagatcctccgccatgggttaaaccatggctaaataaaccaagaaagccagg 910
           | ||  |  || | || |||||| | ||||| ||| |    ||||   | ||||||||
Db   619 CCTCGTCCAGAATTCACCCCCATGGATCAAACCGTGGATCCCAAAAC---GGAAGCCAGG 675

Qy   911 tccccgaatcctggctcttgg---------agagaaaaacaaacactcggccgaaaaagt 961
           | | |||  |||||||||  |          || ||||  |||  |  |  | ||||| |
Db   676 TTCTCGAGCCCTGGCTCTCCGCAGACCAAAAGTGAAACCCAAGGATCCTCCTGAAAAGGA 735

Qy   962 --------------cgagccctcttcctcgtatctaccccgagatcgaggagccgccgac 1007
                        |     |||| ||    ||||| |||||||| || | ||| ||
Db   736 GCCTTTGCCTGGGCCCTCAGCTCTCCCCAAGATCTATCCCGAGATTGAAGGGCCTCCTGA 795

Qy  1008 ttggccggaaccccaacctgttcccccaccccctttatccagcacagggtgctgtgagggg 1067
           ||||| ||||| |  |||       || |  ||   ||||||         ||     |
Db   796 GTGGCCAGAACCTCCTCCTCCCTATCCCCTGCCCCCTCCAGCTTCCACGCCTCCACAGCT 855

Qy  1068 acctctgcccctc-------ctggagctccggtggtggagggacctgctgccgggactcg 1120
           ||   ||||  |       |||| |          || |  ||||| || || ||||| |
Db   856 TCCAGAGCCCGGCCCTGGGGCTGGTGGCGGCACAGTTGGAGGACCGGCGGCTGGGACCCA 915

Qy  1121 gagccggagag---------gcgccaccccggagcggacagacgagatcgcgatattacc 1171
           |||||||| |          | || |    | |     | |||   ||  |    || ||
Db   916 AAGCCGGAGCGCACAGAGTCTCTCCGCAGATGGGGTTCCTGACACCATGTCTGCCTTTCC 975

Qy  1172 gctgcgcacctatggccctcccatgccaggggggccaattgcagcccctccagtattggcc 1231
             || |  |    || ||  |          |  |    | ||| | || ||||||||||
Db   976 TTTGTGTGCAGTGGGACCAGCGGCAGACCTGCACTCTCTCCAGACTCTTCAGTATTGGCC 1035

Qy  1232 cttttcttctgcagatctctataattggaaaactaaccatccccctttctcggaggatcc 1291
           || || || || |||||||||||| |||||        ||||  |||| || ||  ||||
```

```
Db   1036 ATTCTCATCAGCTGATCTCTATAACTGGAA--------ATCCTACTTTTTCAGAAAATCC 1087

Qy   1292 ccaacgcctcacggggttggtggagtcccttatgttctctcaccagcctacttgggatga 1351
          | || ||  |  |||||||||||  | ||||| ||||| ||||| |||| ||||||
Db   1088 AGCTGGACTAACCAGCCTGGTGGAGTCATTAATGTTATCTCATCAGCCCACTTAGGATGA 1147

Qy   1352 ttgtcaacagctgctgcagacactcttcacaaccgaggagcgagagagaattctgttaga 1411
          ||| |||||||| |||| |  || || || || || ||  ||| ||||| || |||
Db   1148 TTGCCAACAGCTTTTGCATGTTCTTTTTACCACAGAAGAAAAAGAAAGAATCCTCTTAAG 1207

Qy   1412 ggctagaaaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatgagattga 1471
          ||||  || | ||| | ||| |  || ||  || ||  |  ||  ||  ||||  |
Db   1208 GGCTCAGAAGACTGTCTCAGGGACAAACAGGGCCCCAACCAACCTGTCCAACAAGATAAA 1267

Qy   1472 catgggatttcccttgactcgccccggttgggactacaacacggctgaaggtagggagag 1531
          | |||| || ||  |  | | || |  ||||||||||||||||||| ||||||||||||||
Db   1268 CGTGGGGTTCCCTCTCGATAGACCAGACTGGGACTACAACACGGCTCAAGGTAGGGAGAG 1327

Qy   1532 cttgaaaatctatcgccaggctctggtggcgggtctccggggcgcctcaagacggcccac 1591
          |||| | | ||  ||| |||||| |||||  |||||  ||| ||  |||||  | ||| |
Db   1328 GTTGACAGTTTACTGCCGGGCTCTAGTGGCAGGTCTCAAGGGAGCGGCAAGATGCCCCTC 1387

Qy   1592 taatttggctaaggtaagagaggtgatgcagggaccgaacgaacctccctcggtatttct 1651
          ||| |||| |||||||||||  |  |||||||| ||    ||| | || || || |||||
Db   1388 CAATCTGGCAAAGGTAAGAGAAGCTATGCAGGGGCCAGCTGAAACCCCGTCTGTTTTTCT 1447

Qy   1652 tgagaggctcatggaagccttcaggcggttcacccctttgatcctacctcagaggccca 1711
          |||||  | |||||  || | | | || | ||| || |||||  || |||||  ||||  ||
Db   1448 AGAGAGATTAATGGAGGCTTACGGACGCTACACTCCCTTTGACCCGACCTCGGAGGGTCA 1507

Qy   1712 gaaagcctcagtggccctggccttcattgggcagtcggctctggatatcaggaagaaact 1771
          |||||  | || ||  |||| || || ||||||||        || || |  |  ||  |
Db   1508 ACAAGCCGCGGTAGCTATGGCTTTTATAGGGCAGTCATTAGCAGACATTAAAAGCAAGTT 1567

Qy   1772 tcagagactggaagggttacaggaggctgagttacgtgatctagtgagagaggcagagaa 1831
          |||||||||||||||||||  |      |  ||||  |     | | ||||||||||||||
Db   1568 ACAGAGACTGGAAGGGTTACATGCCATGGCTTTACAGG----ACTTAGAGAGGCAGAGAA 1623

Qy   1832 ggtgtattacagaagggagacagaagaggagaaggaacagagaaaagaaaaggagagaga 1891
          ||||  |   || || || |||||  |  || ||| || ||||| ||| |   |||  | |
Db   1624 AGTGTTTACAAGGAGAGAAACAGAGGGAGAAAAGAAAGAGAGAGAAGCAT--GAGGAATA 1681

Qy   1892 agaaagggaggaaagacgtgatagacggcaagagaagaatttgactaagatcttggccgc 1951
          |  || | |  | ||  | |||    |||| |||  ||||  |  | |||| |||| ||
Db   1682 AAGAAAGAAAAATAGGAGGGATCAGAGGCAGGAGAGAAATTTAAGCAGGATCCTGGCTGC 1741

Qy   1952 agtggttgaagggaagagcagcagggagagagagagagattttaggaaaattaggtcagg 2011
          ||| || |||| |       | || |  |||       |      | |         ||
Db   1742 AGTAGTAGAAGAGCGTGAGGGAAGAGGTAGATCTGAGAGTAGACAAACAGGGTACCTGGG 1801

Qy   2012 c---------cctagacagtcagggaacctgggcaatagga-cccc-actcgacaaggacca 2062
          |         |||||||  |  || |       |  |  ||   | || || ||||||||
Db   1802 CAACAGGGAACCTAGACCTTGTGGAAGAAAACCCGTTCAGAAAACTCTGGAAAAGGACCA 1861

Qy   2063 gtgtgcgtattgtaaagaaaaaggacactgggcaaggaactgccccaagaagggaaacaa 2122
          | ||| || || ||||| || ||||| |||||  |  | |||||     |  |      ||
Db   1862 ATATGCATAGTGCAAAGAGAAGGGACATTGGGCGCGAGAATGCCCATGAAGAGTGGGAAA 1921

Qy   2123 aggaccgaaggtcctagctctagaagaagataaagattaggggagacggggttcggaccc 2182
          |  || ||||||  |  |||| |||||  |||   || |||||||  | |||| || ||||
Db   1922 TGCCCCTAAGGTCATGACTCTCGAAGATGAT---GACTAGGGGAAATGGGGCTCAGACCT 1978

Qy   2183 cctccccgagcccagggtaactttgaaggt-ggaggggcaaccagttgagttcctggttg 2241
          |||||||  |||| ||||||||  |||  || |||||||   ||  |  || || || ||
Db   1979 CCTCCCCAAGCCTAGGGTAACCCTGACTGTGGGAGGGGACTCCTATGGACTTTCTAGTAT 2038

Qy   2242 ataccggagcggagcattcagtgctgctacaaccattaggaaaact 2287
          |||||||  |  || || || |||  ||  ||||||||| || || ||
Db   2039 ATACCGGGGTAGAACACTCTGTGTTGAAACAACCATTGGGGAAGCT 2084
```

RESULT 2
BI184013

```
LOCUS       BI184013      857 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-cf-b-12-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-cf-b-12-0-UNL 3', mRNA sequence.
ACCESSION   BI184013
VERSION     BI184013.1  GI:14658422
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 857)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .857
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-cf-b-12-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT      152 a    217 c    248 g    236 t      4 others
ORIGIN
```

```
  Query Match             7.2%;  Score 594.6;  DB 11;  Length 857;
  Best Local Similarity   85.6%;  Pred. No. 3.3e-110;
  Matches  773;  Conservative    0;  Mismatches   51;  Indels   79;  Gaps    7;

Qy       2 tggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatca 61
            |||  |||||||||||||||||| | |||||||||||||||||||||||||||||||||
Db      12 TGGCTTACGACTGTGGGCCCCAGTGTGCTTGGAATAAAAATCCTCTTGCTGTTTGCATCA 71

Qy      62 agaccgcttctcgtgagtgatttggggtgtcgcctcttccgagcccggacgagggggatt 121
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      72 AGACCGCTTCTCGTGAGTGATTTGGGGTGTCGCCTCTTCCGAGTCCGGACGAGGGGGATT 131

Qy     122 gttcttttactggcctttcatttggtgcgttggccgggaaatcctgcgaccacccccttac 181
           | ||| | ||||||||||||||||||||||||||||||||| || ||||||||||| ||
Db     132 TTCCTTCTGCTGGCCTTTCATTTGGTGCGTTGGCCGGGAAACCCCGCGACCACCCCTCAC 191

Qy     182 acccgagaaccgacttggaggtaaagggatccccctttggaacgtgtgtgt--gtgtcggc 239
           |||||||||||||||||||||||||||||||||||||||||||||||| |||||||  ||| ||||
Db     192 ACCCGAGAACCGACTTGGAGGTAAAGGGATCCCCTTTGGAATGTGTGTGTGCGTGCCGGC 251

Qy     240 cggcgtctctgttctgagtgtctgttttcggtgatgcgcgctttcggtttgcagctgtcc 299
```

```
          |||||||||||||||||||| ||||||| | ||||||||||||||||||||||||||||
Db    252 CGGCGTCTCTGTTCTGAGTGTTTGTTTTCTGGGATGCGCGCTTTCGGTTTGCAGCTGTCC 311

Qy    300 tctcagaccgtaaggactggaggactgtgatcagcagacgtgctaggaggatcacaggct 359
          || |||||||||||||||| |||||||||||| |||||||||||||||||||||||||||
Db    312 TCCCAGACCGTAAGGACTGGGGGACTGTGATCAGTAGACGTGCTAGGAGGATCACAGGCT 371

Qy    360 gccaccctgggggacgccccgggaggtggggagagccagggacgcctggtggtctcctac 419
          ||| |||||||||||||||||||||||||||||||||    |||||||||||||||| |
Db    372 GCCGCCCTGGGGGACGCCCCGGGAGGTGGGGAGAGCCAGG--TGCCTGGTGGTCTCCTTC 429

Qy    420 tgtcggtcagaggaccgagttctgttgttgaagcgaaagcttccccctccgcggccgtcc 479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 TGTCGGTCAGAGGACCGAGTTCTGTTGTTGAAGCGAAAGCTTCCCCCTCCGCGGCCGTCT 489

Qy    480 gactcttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggtt 539
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    490 GACTCTTTTGCCTGCTTGTGGAAGGCGCGGACGGGTCGCGTGTGTCTGGATCTGTTGGTT 549

Qy    540 tctgtctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagaca 599
          ||||| |||||||||||||||| |||||||||||  ||||||||||||||||||||||
Db    550 TCTGTTTCGTGTGTCTTTGTCTTATGCGTCCTTGTC-CCAGTTTTAATATGGGACAGACG 608

Qy    600 gtgacta------ccccccttagtttgactctcgacca-ttggactgaagttagatccag 652
          ||||| |      |||||||||||||||||||| |||| ||||||||||||||| ||||||
Db    609 GTGACGACCCCCCCCCCCCTTAGTTTGACTCTCAACCATTTGGACTGAAGTTAAATCCAG 668

Qy    653 ggctcataatttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctga 712
          ||||||||||||||||||||||||||||||
Db    669 GGCTCATAATTTGTCAGTTCAGGTTAAGAA---------------------------- 698

Qy    713 atggccaacattcgatgttggatggccatcagaggggaccttttaattctgaaattatcct 772
                                        |||||||||||||||| ||||||||
Db    699 ----------------------------------GGGACCTTTAATTCTGAGATTATCCT 724

Qy    773 ggctgttaaggcaatcatttttcagactggacccggctctcatcctgatcaggagcccta 832
          |||||| || ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    725 GGCTGTNAAAGCAATTATTTTTCAGACTGGACCCGGCTCTCATCCTGATCAGGAGCCCTA 784

Qy    833 tatccttacgtggcaagatttggcagaagatcctccgccatgggttaaaccatggctaaa 892
          ||||||||| |||||||||| |    |     |  |  ||||| ||||||||||||||||
Db    785 TATCCTTACATGGCAAGATTGG---CAGAGATCCTCNGCATGGATTAAACCATGGCTAAT 841

Qy    893 taa 895
            ||
Db    842 AAA 844
```

```
RESULT   3
AK011989
LOCUS       AK011989     2096 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus 10 days embryo cDNA, RIKEN full-length enriched
            library, clone:2610305J24, full insert sequence.
ACCESSION   AK011989
VERSION     AK011989.1  GI:12848454
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2610305J24.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2096)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 2096)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 2096)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 2096)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2096)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .2096
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1905349"
                     /db_xref="MGD:MGI:1919770"
                     /clone="2610305J24"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days embryo"
     CDS             688. .>2096
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB27961.1"
                     /db_xref="GI:12848455"
                     /translation="MGQTVSTPLSLTKDHWTDVRARGQNLSVKVKKKPWMTFCSSEWP
                     VFGVGWPAEGTFYLPTIRAVKAIVFQEGPGSHPDQQPYIMVWEDLARYPPPWVRPFLP
                     PLRPGTKILAIQENGEKEKPKPPLGRDDDRSTPVTKPPKIYSEIEEPPEWPDTPQPPP
                     YAPQPQPSAPSGPLPQAPAGGGGPSTGTRSRRGVTPEGPADSTVALPLRAIGAPPADP
                     NSLQPLQYWPFSSSDLYNWKAHHPPFSENPAGLTGLVESLMYSHQPTWDDCQQLLQTL
                     FTTEERERILLEARKNVRDEAGRPVQTPAEIDEGFPLTRPRWDYNTASGRERLSNYRR
                     VLVAGLRGAARQPTNLAKVREVMQGATEPPSVFLERLMEAYRRYTPFDPTSEGQRASV
                     IMAFIGQSAPDIRKKLQRIEGLQDYTIRDVVREAEKVYHRRETEDKKLEREKREKREE
                     EDRRDRRQEKVLTRILAAVG"
BASE COUNT      500 a    527 c    577 g    492 t
ORIGIN
```

```
Query Match             6.7%;  Score 552.4;  DB 12;  Length 2096;
Best Local Similarity   63.5%;  Pred. No. 1.1e-101;
```

```
  Matches  962;  Conservative    0;  Mismatches  506;  Indels   46;  Gaps    6;

Qy     484 cttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggtttctg 543
           | |||| ||| ||||| ||| |  |     | || |||||  ||||| | || || | |
Db     587 CGTTTGGCTGGTTGTGTAAGTCCAGACGCAGACGAGTGTGCTTGGATGTCTTAGTGTTTT 646

Qy     544 tctcgtgtgtctttgtcttgtgcgtccttgtctacagttttaatatgggacagacagtga 603
           |   | |||| |||| |||||  | ||  | | |   || | ||||||||||| |||
Db     647 CCGTCTCTGTCATTGTGTTGTGTTTACTCTTATTC---TTCACTATGGGACAGACCGTGT 703

Qy     604 ctacccccctagtttgactctcgaccattggactgaagttagatccagggctcataatt 663
           |||| ||  |   |||||||  |||||||||| || |||||  | || |  || ||||
Db     704 CTACTCCTTTATCTTTGACTAAGGACCATTGGACGGACGTTAGGGCTAGAGGACAAAATT 763

Qy     664 tgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggccaacat 723
           |||||||  | || |||||    || |||  |||||||||| |||| ||||||||    |
Db     764 TGTCAGTAAAAGTGAAGAAAAAGCCATGGATGACTTTCTGTTCCTCAGAATGGCCTGTTT 823

Qy     724 tcgatgttggatggccatcagagggaccttttaattctgaaattatcctggctgttaagg 783
           | |  || || |||||| |||| || || ||| | |      |  ||   ||||| ||||
Db     824 TTGGAGTAGGTTGGCCAGCAGAAGGAACTTTTTACTTACCCACCATAAGGGCTGTGAAGG 883

Qy     784 caatcatttttcagactggacccggctctcatcctgatcaggagccctatatccttacgt 843
           | ||  ||||||||   ||||| || || ||||| || ||  | || || ||| |    |
Db     884 CCATTGTTTTTCAGGAAGGACCAGGGTCGCATCCAGACCAACAACCGTACATCATGGTAT 943

Qy     844 ggcaagatttggcagaagatcctccgccatgggttaaaccat------ggctaaataaac 897
           || | || ||||||    | || || || ||||||   ||||      ||        |
Db     944 GGGAGGACTTGGCACGCTACCCACCCCCGTGGGTTCGCCCATTCCTCCCGCCTCTCCGCC 1003

Qy     898 caagaaagccaggtccccgaatcctggctcttggagagaaaaacaaac------------ 945
           | | |        ||     ||||  |    ||| |||||| | ||||
Db    1004 CTGGCACCAAGATTCTAGCCATCCAAGAAAATGGTGAGAAAGAGAAACCGAAACCACCGC 1063

Qy     946 --------------actcggccgaaaaagtcgagccctcttcctcgtatctaccccgaga 991
                          |||    | |     ||     |  ||    |||||  | ||||
Db    1064 TCGGGAGAGATGATGATCGCAGCACACCAGTGACGAAACCCCCCAAGATCTATTCAGAGA 1123

Qy     992 tcgaggagccgccgacttggccgga---accccaacctgttccccccacccccttatccag 1048
           | || || || ||    |||||  ||    || |||||     ||      | ||  | ||
Db    1124 TTGAAGAACCCCCTGAGTGGCCCGATACCCCTCAACCCCCACCTTATGCTCCCCAGCCCC 1183

Qy    1049 cacagggtgctgtgaggggacctctgcccctcctggagctccggtggtggagggacctgc 1108
           ||    |||     |||||| |||  ||||  |    |  | |  |  | ||| ||  |
Db    1184 AACCTTCAGCTCCCTCGGGACCCCTG--CCTCAGGCCCCGGCCGGAGGAGGGGGTCCCTC 1241

Qy    1109 tgccgggactcggagccggagaggcgccaccccggagcggacagacgagatcgcgatatt 1168
             | ||  ||  |||||||| |||| | |||||| ||| || | |  ||      |  |
Db    1242 CACAGGAACAAGGAGCCGGCGAGGAGTCACCCCTGAGGGGCCTGCGGATTCAACTGTGGC 1301

Qy    1169 accgctgcgcacctatggccctcccatgccaggggggccaa------ttgcagcccctcca 1222
           |  |  | ||    |       |    || |  |  | |       | ||||||| ||
Db    1302 GCTCCCCCTCAGGGCTATTGGGGCTCCCCCTGCCGATCCAAATAGTCTACAGCCCCTACA 1361

Qy    1223 gtattggccctttttcttctgcagatctctataattggaaaactaaccatccccctttctc 1282
           ||||||||| ||||| ||| | || || || || |||||| || | || |||||||| ||
Db    1362 GTATTGGCCTTTTTCCTCTTCTGACCTTTACAACTGGAAAGCTCATCACCCCCCTTTTTC 1421

Qy    1283 ggaggatccccaacgcctcacggggttggtggagtcccttatgttctctcaccagcctac 1342
             ||  | ||   | | ||||| |||||||| || ||  | ||||  || ||||||| ||
Db    1422 AGAAAACCCTGCAGGACTCACTGGGTTGGTTGAATCATTAATGTATTCACACCAGCCGAC 1481

Qy    1343 ttgggatgattgtcaacagctgctgcagacactcttcacaaccgaggagcgagagagaat 1402
           |||||||| || || ||||| |||||||| || ||||||||||||||| |||||||  ||
Db    1482 CTGGGATGACTGCCAGCAGCTTCTGCAGACTCTATTCACAACCGAGGAGAGAGAGAGGAT 1541

Qy    1403 tctgttagaggctagaaaaaatgttcctggggccgacgggcgacccacgcagttgcaaaa 1462
           |||  | |||||| | |||||||| |  |  |  |  ||||| ||    ||   | |
Db    1542 TCTCCTCGAGGCTCGGAAAAATGTCCGAGACGAGGCTGGGCGCCCTGTCCAAACTCCAGC 1601

Qy    1463 tgagattgacatgggatttcccttgactcgccccggttgggactacaacacggctgaagg 1522
           |||||| ||    ||||||||  | || || ||| | ||||| || || |||||   |||
Db    1602 TGAGATAGATGAAGGATTTCCGCTAACCCGGCCCCGATGGGATTATAATACGGCATCAGG 1661

Qy    1523 tagggagagcttgaaaatctatcgccaggctctggtggcgggtctccggggcgcctcaag 1582
           ||||||  |  ||   |  ||||||| ||  || |||||||||||| | || ||  |  |
Db    1662 TAGGGAACGACTGTCCAATTATCGCCGGGTCCTAGTGGCGGGTCTCAGAGGTGCTGCCCG 1721

Qy    1583 acggcccactaatttggctaaggtaagagaggtgatgcagggaccgaacgaacctccctc 1642
           | |||||| ||| |||| |||||||||||||| ||||||||| |||  || || |||||
Db    1722 GCAGCCCACGAATCTGGCCAAGGTAAGAGAGGTTATGCAGGGAGCGACTGAGCCCCCCTC 1781

Qy    1643 ggtatttcttgagaggctcatggaagccttcaggcggttcacccctttgatcctacctc 1702
           || || ||||| |||||||||| || |  ||| | |  ||||| || || || || ||
Db    1782 AGTCTTCCTTGAAAGGCTCATGGAGGCTTATAGGAGATATACCCCATTCGACCCCACGTC 1841

Qy    1703 agaggcccagaaagcctcagtggccctggccttcattgggcagtcggctctggatatcag 1762
           ||||  || |  ||||||||     |||||||||||| ||||| ||||  || || ||
Db    1842 TGAGGGTCAAAGGGCCTCAGTAATTATGGCCTTCATTGGCCAGTCAGCTCCTGACATTAG 1901

Qy    1763 gaagaaacttcagagactggaagggttacaggaggctgagttacgtgatctagtgagaga 1822
           ||||||  | ||| || | || || || |||||         || | ||| |||| |||||
Db    1902 GAAGAAGTTACAGCGAATTGAGGGCTTGCAGGATTACACCATAAGGGATGTAGTTAGAGA 1961

Qy    1823 ggcagagaaggtgtattacagaagggagacagaagaggagaaggaacagagaaaagaaaa 1882
           |||||||||  |||||| | || || || ||||||||  | |||  | |||||  |  |||
Db    1962 GGCAGAGAAAGTGTATCATAGGAGAGAAACAGAAGATAAAAAGTTAGAGAGAGAGAAAAG 2021

Qy    1883 ggagagagaagaaagggaggaaagacgtgatagacggcaagagaagaatttgactaagat 1942
           |||| |  ||||  |||||| ||  | || ||  ||||||| |||  |||||||| |||
Db    2022 AGAGAAAAGAGAAGAGGAGGATAGGAGAGACAGGAGGCAAGAAAAGGTTTTGACTAGGAT 2081

Qy    1943 cttggccgcagtgg 1956
           | |||||||||| |
Db    2082 CCTGGCCGCAGTAG 2095
```

```
RESULT   4
AW657531
LOCUS       AW657531      549 bp    mRNA            EST       25-APR-2001
DEFINITION  110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW657531
VERSION     AW657531.1  GI:7423429
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 94  row: E  column: 12
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Bos taurus"
```

```
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT      145 a    144 c    121 g    139 t
ORIGIN


  Query Match             5.8%;  Score 473.6;  DB 10;  Length 549;
  Best Local Similarity   92.7%;  Pred. No. 1.1e-85;
  Matches  508;  Conservative    0;  Mismatches   39;  Indels    1;  Gaps    1;

Qy   6483 ctccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagggagcttt 6542
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 CTCCCCCGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGGGAGCTTT 61

Qy   6543 ccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtctatcctc 6602
          |||||||||||||||||||||||| ||||||| ||||||||||||||||||||||||||||
Db     62 CCAAGCCATCAACTCCACCGACCCCGATGCCAGTTCTTCTTGTTGGCTTTGTCTATCCTC 121

Qy   6603 agggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagca 6662
          |||||||||||||||||||| ||||||| |||| | ||||||||||||||||||||||||
Db    122 AGGGCCTCCTTATTATGAGGGAATGGCTAGAGAAGGGAAATTCAATGTGACCAAAGAGCA 181

Qy   6663 tagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagtttccgggaa 6722
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    182 TAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTCACCCTCACTGAAGTTTCCGGGAA 241

Qy   6723 ggggacatgcataggaaaagctccccccatcccaccaacaccttttgctatagtactgtggt 6782
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||
Db    242 GGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCAATAGTACTATGGT 301

Qy   6783 ttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggcatg 6842
          |||||||||||||| |||||||||||||||||||||||||||||| | ||||||||||||
Db    302 TTATGAGCAGGCCTTAGAAAATCAGTATTTAGTACCTGGTTATAACTGATGGTGGGCATG 361

Qy   6843 caatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttctg 6902
          |||||||||||||| |||||||||| ||| ||||||| ||||||||||||||||||||||
Db    362 CAATACTGGGTTAACTCCCTGTGTTTCTACCACAGTCTTTAACCAATCCAAAGATTTCTG 421

Qy   6903 tgtcatggtccaaatcgtccccgagtgtactaccatcctgaggaagtggtccttgatga 6962
          |||||||||||| |||||||| | ||  |||||||||||||||||| ||| | |||||
Db    422 TGTCATGGTCCAGCTCGTCCCCTGGGTACACTACCATCCTGAGGAAGTAGTCATCGATGA 481

Qy   6963 atatgactatcggtataaccgaccaaaaagagaacccgtatccctt-accctagctgtaa 7021
          |||||||||||  || |||  |||||||||| |||||  ||||| |||||||| || |
Db    482 GTATGACTATCGGCCTACCCGCTCAAAAAGAGAGCCCGTGACCCTTAACCCTAGCCGTTA 541

Qy   7022 tgctcgga 7029
          | || |||
Db    542 TACTAGGA 549


RESULT   5
AK004927
LOCUS       AK004927     2935 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male liver cDNA, RIKEN full-length enriched
            library, clone:1300007C21, full insert sequence.
ACCESSION   AK004927
VERSION     AK004927.1  GI:12836484
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:1300007C21.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2935)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 2935)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 2935)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 2935)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2935)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES                Location/Qualifiers
     source             1. .2935
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /db_xref="MGD:MGI:1897042"
                        /db_xref="MGD:MGI:1914777"
                        /clone="1300007C21"
                        /sex="male"
                        /tissue_type="liver"
```

```
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
BASE COUNT       669 a    860 c    773 g    633 t
ORIGIN
```

Query Match 5.7%; Score 472; DB 12; Length 2935;
Best Local Similarity 54.4%; Pred. No. 2e-85;
Matches 1208; Conservative 0; Mismatches 965; Indels 47; Gaps 11;

```
Qy    485 ttttgcctgcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggtttctgt 544
Db    530 TTTTGCTTTCGGTTTTACGCCGAAACCGCGCCGCGCGTCTGATTTGTTTGTTGTTGTGT 589

Qy    545 ctcgtgtgtctttgtcttgtgcgtccttgtctacagtttt------aatatgggacagac 598
Db    590 TCTTTTGTTCTTCGTTAGTTTTCTCCTGTCTTTAAGTGTTTTCGAGATCATGGGACAGAC 649

Qy    599 agtgacta-ccccccttagtttgactctcgaccattggactgaagttagatccagggctc 657
Db    650 CGTAACTACCCCCCCTGAGTCTAACCTTGCAGCACTGGGGAGATGTCCAGCGCATTGCAT 709

Qy    658 ataatttgtcagttcaggttaagaagggaccttggcagactttctgtgcctctgaatggc 717
Db    710 CCAATCAGTCTGTGGATGTCAGGAAGAGGCGCTGGGTTACTTTCTGTTCCGCCGAATGGC 769

Qy    718 caacattcgatgttggatggccatcagaggggacctttaattctgaaattatcctggctg 777
Db    770 CAACTTTCAATGTGGGATGGCCTCAGGATGGTGCTTTTAATTTGGGTATTATCTCTCAGG 829

Qy    778 ttaaggcaatcatttttcagactggacccggctctcatcctgatcaggagccctatatcc 837
Db    830 TCAAGTCTAGAGTGTTTTGTCCTGGTCCCCACGGACACCCGGATCAGGTCCCATATATCG 889

Qy    838 ttacgtggcaag-atttggcagaagatcctccgccatgggttaaaccatggctaaataaa 896
Db    890 TCACCTGGGAGGCACTCTCCAAAACCCCCTCCTTTACCGACAGCTCCCGTCCTCCCGCCC 949

Qy    897 ccaagaaagccaggtccccgaatcctggctcttggagagaaaaacaaacactcggccgaa 956
Db    950 GGTCCTTCTGCGCAACCTCCGTCCCGATCTGCCCTTTACCCCGCCCTTACCCCCTCTATA 1009

Qy    957 aaagtcgagccctcttcctcgtatctaccccgagatcgaggagccgccgacttggccgga 1016
Db   1010 AAGCCCAAACCTCCTAAGCCCCAGGTTCTCCCTGATAGCGGCGGACCTCTCAT----TGA 1065

Qy   1017 accccaacctgttcccccaccccctatccagcacagggtgctgtgaggggacctctgcc 1076
Db   1066 CCTTCTCACAGAGGACCCCCCGCCGTACGGAGCACAACCTTCCTCCTCTGCCAGAGAGGA 1125

Qy   1077 cctcctggagctccggtggtggagggacctgctgccgggactcggagccggagaggcgcc 1136
Db   1126 CGATGAAAAAGAGGCGGCCGCCACCTCCGAGGTTCCCCCCCCTTCTCCCATGGTGTCTCG 1185

Qy   1137 accccggagcggacagacgagatcgcgatattaccgctgcgcacctatggccctcccatg 1196
Db   1186 ACTGCGGGGAAGGAGGGACCCTCCCGCAGCGGACTCCACCTCCTCCCAGGCATTCCCACT 1245

Qy   1197 ccaggggccaattgcagcc-cctccagtattggccctttttcttctgcagatctctataa 1255
Db   1246 CCGTATGGGGGGAGATGGCCAGCTTCAGTACTGGCCGTTTTCCTCCTCTGACTT------ 1299

Qy   1256 ttggaaaactaaccatccccctttctcggaggatccccaacgcctcacggggttggtgga 1315
Db   1300 ----AAAAATAATAACCCTTCCTTTTCTGAAGACCCAGGTAAATTGACGGCTTTGATTGA 1355

Qy   1316 gtcccttatgttctctcaccagcctacttgggatgattgtcaacagctgctgcagacact 1375
Db   1356 GTCCGTCCTCATCACCCACCAGCCCACCTGGGACGACTGTCAGCAGTTGTTGGGGACCTT 1415

Qy   1376 cttcacaaccgaggagcgagagagaattctgttagaggctagaaaaaatgttcctggggc 1435
Db   1416 GCTGACCGGAGAAGAAAAGCAGCGGGTGCTCCTAGAGGCTAGAAAGGCAGTCCGGGGCAA 1475

Qy   1436 cgacgggcgacccacgcagttgcaaaatgagattgacatgggatttcccttgactcgccc 1495
Db   1476 TGATGGACGCCCCACTCAGTTGCCTAATGAAGTCAATGCTGCTTTTCCCCTTGAACGCCC 1535

Qy   1496 cggttgggactacaacacggctgaaggtagggagagcttgaaaatctatcgccaggctct 1555
Db   1536 CGATTGGGATTACACCACTACAGAAGGTAGGAACCACCTAGTCCTCTATCGCCAGTTGCT 1595

Qy   1556 ggtggcgggtctccggggcgcctcaagacggcccactaatttggctaaggtaagagaggt 1615
Db   1596 CTTAGCGGGTCTCCAAAACGCGGGCAGAAGCCCCACCAATTTGGCCAAGGTAAAAGGGAT 1655

Qy   1616 gatgcagggaccgaacgaacctccctcggtatttcttgagaggctcatggaagccttcag 1675
Db   1656 AACCCAGGGACCTAATGAGTCTCCCTCAGCCTTTTTAGAGAGACTCAAGGAGGCATATCG 1715

Qy   1676 gcggttcacccctttttgatcctacctcagaggcccagaaagcctcagtggccctggcctt 1735
Db   1716 CAGGTACACTCCTTATGACCCTGAGGACCCAGGGCAAGAAACCAATGTGTCTATGTCATT 1775

Qy   1736 cattgggcagtcggctctggatatcaggaagaaacttcagagactggaagggttacagga 1795
Db   1776 CATCTGGCAGTCTGCCCCGGATATCGGGCGAAAGTTAGAGCGGTTAAAAGATTTAAAGAG 1835

Qy   1796 ggctgagttacgtgatctagtgagagaggcagagaaggtgtattacagaagggagacaga 1855
Db   1836 CAAGACCTTAGGAGACTTAGTGAGGGAAGCTGAAAAGATCTTTAATAAGCGAGAAACCCC 1895

Qy   1856 agaggagaaggaacagagaaaagaaaaggagagagaagaaaggggaggaaagacgtgatag 1915
Db   1896 GGAAGAAAGAGAGGAACGTATCAGGAGAGAAACAGAGGAAAAGGAAGAACGCCGTAGGGC 1955

Qy   1916 acggcaagagaagaatttgactaagatcttggccgcagtggttgaagggaagagcagcag 1975
Db   1956 AGAGGATGAGCAGAGAGAGAAAGAAAGGGACCGCAGGAGACATAGAGAGATGAGCAAGCT 2015

Qy   1976 ggagagagagagagattttaggaaaattaggtcaggccctagacagtcagggaacctggg 2035
Db   2016 CTTGGCCACTGTAGTTAGTGGACAGAGACAGGGTAGACAG---------GGGGGAGAGCG 2066

Qy   2036 caataggaccccactcgacaaggaccagtgtgcgtattgtaaagaaaaaggacactgggc 2095
Db   2067 AAGGAGGCCCCAACTTGATAAGGACCAATGCGCCTACTGCAAAGAAAAAGGACACTGGGC 2126

Qy   2096 aaggaactgccccaagaag------ggaaacaaaggaccgaaggtcc---tagctctaga 2146
Db   2127 TAAGGACTGCCCAAAGAAGCCACGAGGACCCCGAGGACCGAGGCCCCAGACCTCCCTCCT 2186

Qy   2147 agaagataaagattaggggagacggggttcggacccccctccccgagcccagggtaacttt 2206
Db   2187 AACCTTAGGTGACTAGGGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCT 2246

Qy   2207 gaaggtggaggggcaaccagttgagttcctggttgataccggagcggagcattcagtgct 2266
Db   2247 CAAAGTCGGGGGGCAACCCGTCACCTTCCTGGTAGATACTGGGGCCCAACACTCCGTGCT 2306

Qy   2267 gctacaaccattaggaaaactaaaagaaaaaaaatcctgggtgatgggtgccacagggca 2326
Db   2307 GACCCAAAATCCTGGACCCCTAAGTGACAAGTCTGCCTGGGTCCAAGGGGCTACTGGAGG 2366

Qy   2327 acggcagtatccatggactacccgaagaaccgttgacttgggagtgggacgggtaaccca 2386
Db   2367 AAAGCGGTATCGCTGGACCACGGATCGCAAAGTACATCTAGCTACCGGTAAGGTCACCCA 2426

Qy   2387 ctcgtttctggtcatccctgagtgcccagtacccctctaggtagagacttactgaccaa 2446
Db   2427 CTCTTTCCTCCATGTACCAGACTGCCCCTATCCTCTGCTAGGAAGAGACTTGCTGACTAA 2486

Qy   2447 gatgggagctcaaatttcttttga---acaaggaagaccagaagtgtctgtgaataacaa 2503
Db   2487 ACTCAAAGCCCAGATCCACTTCGAGGGATCAGGAGCTCAGGTTGTGGGACCAATGGGACA 2546
```

```
Qy    2504 acccatcactgtgttgaccctccaattagatgatgaatatcgactatattctccccaagt 2563
           |||  |    ||| |||||||  |  |||| |||||  |||||  |||  ||    ||  |
Db    2547 GCCCCTGCAAGTGCTGACCCTAAACATAGAAGATGAGTATCGGCTACATGAGACCTCAAC 2606

Qy    2564 aaagcctgatcaagatatacagt---cctggttggagcagtttccccaagcctgggcaga 2620
           | ||||  |||     |  ||   ||   |||||  |     |  ||||||||  |||||||| ||
Db    2607 AGAGCCGGATGTTTCTCTAGGGTCCACCTGGCTTTCTGATTTTCCCCAGGCCTGGGCGGA 2666

Qy    2621 aaccgcagggatgggtttggcaaagcaagttcccccacaggttattcaactgaaggccag 2680
           |||  |   ||  |||||  |||||    |       |  ||  |  |  ||||  |  ||||||||  ||
Db    2667 AACTGGGGGCATGGGACTGGCAGTTCGCCAAGCGCCTCTGATTATACCTCTGAAGGCAAG 2726


RESULT   6
BF441465
LOCUS       BF441465      467 bp    mRNA            EST       01-DEC-2000
DEFINITION  257856 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF441465
VERSION     BF441465.1  GI:11501629
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 86  row: C  column: 8
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      115 a    118 c    127 g    107 t
ORIGIN


  Query Match             5.1%;  Score 420.6;  DB 11;  Length 467;
  Best Local Similarity   93.8%;  Pred. No. 6.4e-75;
  Matches  438;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy    1299 ctcacgggggttggtggagtcccttatgttctctcaccagcctacttgggatgattgtcaa 1358
           |||||||  ||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTCACGGCGTTAGTGGAGTCCCTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAG 60

Qy    1359 cagctgctgcagacactcttcacaaccgaggagcgagagagaattctgttagaggctaga 1418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CAGCTGCTGCAGACACTCTTCACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAGA 120

Qy    1419 aaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatgagattgacatggga 1478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     121 AAAAATGTTCCTGGGGCCGACGGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGA 180

Qy    1479 tttcccttgactcgccccggttgggactacaacacggctgaaggtagggagagcttgaaa 1538
           |||||||||||||||||||| ||||||||||||||||  ||||||||||||||||||||
Db     181 TTTCCCTTGACTCGCCCCGGATGGGACTACAACACGGCCGAAGGTAGGGAGAGCTTGAAA 240

Qy    1539 atctatcgccaggctctggtggcgggtctccggggcgcctcaagacggcccactaatttg 1598
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     241 ATCTATCGCCAGGCTCTGGTGGCGGGTCTCCGAGGCGCCTCAAGACGGCCCACTAATTTG 300

Qy    1599 gctaaggtaagagaggtgatgcagggaccgaacgaacctccctcggtatttcttgagagg 1658
           ||||||||||||||  |||||||||||| ||| |||||  |||||  || ||||||||||||
Db     301 GCTAAGGTAAGAGAAATGATGCAGGGACTGAATGAACCCCCCTCTGTTTTTCTTGAGAGG 360

Qy    1659 ctcatggaagccttcaggcggttcacccctttgatcctacctcagaggcccagaaagcc 1718
           |||  |||||||||||||||||  ||||||||||||||  ||||||||||||  | ||||||
Db     361 CTCTTGGAAGCCTTCAGGCGGTACACCCCTTTTGATCCCACCTCAGAGGCCTAAAAAGCC 420

Qy    1719 tcagtggccctggccttcattgggcagtcggctctggatatcaggaa 1765
           ||||||||  |||||||  ||  ||  |||||.||  |||||||  |  .||
Db     421 TCAGTGGCTTTGGCCTTTATAGGACAGTCAGCCTTGGATATTAAAAA 467


RESULT   7
BF441468
LOCUS       BF441468      466 bp    mRNA            EST       01-DEC-2000
DEFINITION  257862 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF441468
VERSION     BF441468.1  GI:11501632
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 86  row: D  column: 8
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      114 a    118 c    127 g    107 t
ORIGIN


  Query Match             5.1%;  Score 419.8;  DB 11;  Length 466;
  Best Local Similarity   94.2%;  Pred. No. 9.3e-75;
  Matches  436;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;
```

```
Qy   1299 ctcacggggttggtggagtcccttatgttctctcaccagcctacttgggatgattgtcaa 1358
          ||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTCACGGCGTTAGTGGAGTCCCTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAG 60

Qy   1359 cagctgctgcagacactcttcacaaccgaggagcgagagagaaattctgttagaggctaga 1418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGCTGCTGCAGACACTCTTCACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAGA 120

Qy   1419 aaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatgagattgacatggga 1478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAAAATGTTCCTGGGGCCGACGGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGA 180

Qy   1479 tttcccttgactcgccccggttgggactacaacacggctgaaggtagggagagcttgaaa 1538
          |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||
Db    181 TTTCCCTTGACTCGCCCCGGATGGGACTACAACACGGCCGAAGGTAGGGAGAGCTTGAAA 240

Qy   1539 atctatcgccaggctctggtggcgggtctccggggcgcctcaagacggcccactaatttg 1598
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    241 ATCTATCGCCAGGCTCTGGTGGCGGGTCTCCGAGGCGCCTCAAGACGGCCCACTAATTTG 300

Qy   1599 gctaaggtaagagaggtgatgcagggaccgaacgaacctccctcggtatttcttgagagg 1658
          |||||||||||||  |||||||||||| ||| ||||| |||| || ||||||||||||
Db    301 GCTAAGGTAAGAGAAATGATGCAGGGACTGAATGAACCCCCCTCTGTTTTTCTTGAGAGG 360

Qy   1659 ctcatggaagccttcaggcggttcacccctttgatcctacctcagaggcccagaaagcc 1718
          ||| |||||||||||||||| |||||||||||||| |||||||||||| | ||||||
Db    361 CTCTTGGAAGCCTTCAGGCGGTACACCCCTTTTGATCCCACCTCAGAGGCCTAAAAAGCC 420

Qy   1719 tcagtggccctggccttcattgggcagtcggctctggatatca 1761
          |||||||  ||||||| || || ||||| ||  ||||||| |
Db    421 TCAGTGGCTTTGGCCTTTATAGGACAGTCAGCCTTGGATATTA 463


RESULT   8
BI399234/c
LOCUS       BI399234      465 bp    mRNA            EST       14-AUG-2001
DEFINITION  MI-P-AY1-nrp-a-02-0-UI.s1 MI-P-AY1 Sus scrofa cDNA clone
            MI-P-AY1-nrp-a-02-0-UI 3', mRNA sequence.
ACCESSION   BI399234
VERSION     BI399234.1  GI:15178301
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,
            University of Iowa Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-AY1-nrp-a-02-0-UI"
                     /clone_lib="MI-P-AY1"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-AY1
                     library is normalized library derived from the MI-P-AY0
                     library,  ultimately derived from placenta tissue. For a
                     detailed description of the  library from which this clone
                     was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/. The procedure used to
                     create this library has been previously described (Bonaldo
                     , Lennon and Soares, Genome Research 6: 791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT      137 a    122 c    100 g    106 t
ORIGIN


  Query Match             5.1%;  Score 419.4;  DB 11;  Length 465;
  Best Local Similarity   95.4%;  Pred. No. 1.1e-74;
  Matches  432;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy    493 gcttgtggaagacgcggacgggtcgcgtgtgtctggatctgttggtttctgtctcgtgtg 552
          |||||||||||||| ||||||||| |||||||||||||||||||||||| | |||||
Db    465 GCTTGTGGAAGACGTGGACGGGTCACGTGTGTCTGGATCTGTTGGTTTCTGTTTTGTGTG 406

Qy    553 tctttgtcttgtgcgtccttgtctacagttttaatatgggacagacagtgactacccccc 612
          | |||||||||| |||||||||||||||||||||||||||||| ||||| ||||| |
Db    405 TTTTTGTCTTGTGTGTCCTTGTCTACAGTTTTAATATGGGACAGACGGTGACGACCCCTC 346

Qy    613 ttagtttgactctcgaccattggactgaagttagatccagggctcataatttgtcagttc 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 TTAGTTTGACTCTCGACCATTGGACTGAAGTTAAATCCAGGGCTCATAATTTGTCAGTTC 286

Qy    673 aggttaagaagggaccttggcagactttctgtgcctctgaatggccaacattcgatgttg 732
          |||||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db    285 AGGTTAAGAAGGGACCTTGGCAGACTTTCTGTGTCTCTGAATGGCCGACATTCGATGTTG 226

Qy    733 gatggccatcagaggggacctttaattctgaaattatcctggctgttaaggcaatcattt 792
          |||||||||||||||||||||||||||||| |||||||||||||||| ||||| ||||
Db    225 GATGGCCATCAGAGGGGACCTTTAATTCTGAGATTATCCTGGCTGTTAAAGCAATTATTT 166

Qy    793 ttcagactggacccggctctcatcctgatcaggagccctatatccttacgtggcaagatt 852
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    165 TTCAGACTGGACCCGGCTCTCATCCCGATCAGGAGCCCTATATCCTTACGTGGCAAGATT 106

Qy    853 tggcagaagatcctccgccatgggttaaaccatggctaaataaaccaagaaagccaggtc 912
          ||||||| |||||||||||||||||||||||||||| ||||| ||||||||||||||||
Db    105 TGGCAGAGGATCCTCCGCCATGGGTTAAACCATGGCTGAATAAGCCAAGAAAGCCAGGTC 46

Qy    913 cccgaatcctggctcttggagagaaaaacaaac 945
          ||||||| |||||||||||||||||| ||||
Db     45 CCCGAATTCTGGCTCTTGGAGAGAAAAAAAAAC 13


RESULT   9
BH063533
LOCUS       BH063533      801 bp    DNA             GSS       18-JUL-2001
DEFINITION  RPCI-24-370H8.TJ RPCI-24 Mus musculus genomic clone RPCI-24-370H8,
            DNA sequence.
ACCESSION   BH063533
VERSION     BH063533.1  GI:14878283
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-370H8.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
```

```
           9712 Medical Center Dr., Rockville, MD 20850, USA
           Tel: 301 838 0200
           Fax: 301 838 0208
           Email: szhao@tigr.org
           Clones are derived from the mouse BAC library RPCI-24. For BAC
           library availability, please contact Pieter de Jong
           (pdejong@mail.cho.org). Clones may be purchased from BACPAC
           Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
           page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
           Plate: 370  row: H  column: 8
           Seq primer: SP6
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-370H8"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      227 a    193 c    217 g    164 t
ORIGIN


  Query Match             4.9%;  Score 404.2;  DB 13;  Length 801;
  Best Local Similarity   70.7%;  Pred. No. 1.3e-71;
  Matches  552;  Conservative    0;  Mismatches  228;  Indels    1;  Gaps    1;

Qy   1230 ccctttttcttctgcagatctctataattggaaaactaaccatccccctttctcggaggat 1289
           || ||||| ||| | ||| | |||| ||||||||||||||| || || ||||| ||| |
Db      1 CCTTTTTCCTCTTCTGAT-TATATACTTGGAAAACTAACCACCCTCCCTTCTCAGAGAAC 59

Qy   1290 ccccaacgcctcacggggttggtggagtcccttatgttctctcaccagcctacttgggat 1349
           |||    | || || ||| |  | ||||| |||||||||| || || || |||||||||
Db     60 CCCTCTGGGCTTACTGGGCTCCTTGAGTCACTTATGTTCTCCCATCAACCCACTTGGGAT 119

Qy   1350 gattgtcaacagctgctgcagacactcttcacaaccgaggagcgagagagaattctgtta 1409
           |||||||| |||||  |||||   || ||||| || || ||  |||| ||||| ||| |
Db    120 GATTGTCAGCAGCTTTTGCAGGTTCTTTTCACCACAGAAGAAAGAGAAAGAATCCTGATG 179

Qy   1410 gaggctagaaaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatgagatt 1469
           ||||| ||||||||||||  || |  |||||    |||||     | |  ||     |
Db    180 GAGGCGAGAAAAAATGTTCTGGGAGAGGACGGCACTCCCACTGCCCTCCCTAACCTCGTG 239

Qy   1470 gacatgggatttcccttgactcgccccggttgggactacaacacggctgaaggtagggag 1529
           |||  ||  || |||||||  ||||||   |||||||||||||| || ||||||||||
Db    240 GACGAGGCTTTCCCCTTGAACCGCCCCAACTGGGACTACAACACCGCGGAAGGTAGGGGA 299

Qy   1530 agcttgaaaatctatcgccaggctctggtggcgggtctccggggcgcctcaagacggccc 1589
           || |     |||||||| | |||| ||||| |||||| | || ||| | |||||||||
Db    300 CGCCTCCTTGTCTATCGCCGGACTCTAGTGGCAGGTCTCAGAGGAGCCACTAGACGGCCC 359

Qy   1590 actaatttggctaaggtaagagaggtgatgcagggaccgaacgaacctccctcggtattt 1649
           || |||||||||||||||||||||||  ||||||| | ||  ||||| ||||| || ||
Db    360 ACCAATTTGGCTAAGGTAAGAGAGGTCTTGCAGGGGCAGACTGAACCACCCTCAGTCTTC 419

Qy   1650 cttgagaggctcatggaagccttcaggcggttcacccctttgatcctacctcagaggcc 1709
           |||||  | |||||||| || |  ||| ||| |||||||||||| ||  | |||||||
Db    420 CTTGATCGTCTCATGGAGGCATATAGGAGGTACACCCCTTTTGACCCCTCGTCAGAGGGG 479

Qy   1710 cagaaagcctcagtggccctggccttcattgggcagtcggctctggatatcaggaagaaa 1769
           ||||||||| | || ||| |||||||||||||| ||||| ||||  || || | ||| ||
Db    480 CAGAAAGCCGCTGTAGCCATGGCCTTCATTGGTCAGTCCGCTCCCGACATTAAGAAAAAG 539

Qy   1770 cttcagagactggaagggttacaggaggctgagttacgtgatctagtgagagaggcagag 1829
           || || || ||||| ||| | || ||   |  | | |  ||| |||| | ||| ||||||
Db    540 CTGCAAAGGCTGGAGGGGCTCCAAGATCATACGCTCCAAGATTTAGTAAAAGAAGCAGAG 599

Qy   1830 aaggtgtattacagaagggagacagaagaggagaaggaacagagaaaagaaaaggagaga 1889
           || || ||| | |  ||||| ||   ||| |||| | |  ||||| |  | || || |
Db    600 AAAGTCTATCATAAGAGGGAAACCAGAGAAGAGAGGCAGGAGAGAGAGAAGAAAGAAATG 659

Qy   1890 gaagaaagggaggaaagacgtgatagacggcaagagaagaatttgactaagatcttggcc 1949
           || || |||||  | ||||| ||| | || || |||| |||||||| ||| || ||||||
Db    660 GAGGAGAGGGAAAATAGACGGGATCGGCGTCAGGAGAGAAATTTGAGTAAAATTTTGGCC 719

Qy   1950 gcagtggttgaagggaagagcagcagggagagagagagagattttaggaaaattaggtca 2009
           ||||| ||  | | |        || |  |  | | ||   ||  |   |  ||| ||
Db    720 GCAGTTGTGAATGATAGACAGTCAGGGAAAGGTAAAACAGGACTTCTGGGCAACAGGGCA 779

Qy   2010 g 2010
           |
Db    780 G 780


RESULT  10
BF441466
LOCUS       BF441466      452 bp    mRNA            EST       01-DEC-2000
DEFINITION  257857 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF441466
VERSION     BF441466.1  GI:11501630
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 86  row: C  column: 9
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      107 a    118 c    125 g    102 t
ORIGIN


  Query Match             4.9%;  Score 401.6;  DB 11;  Length 452;
  Best Local Similarity   94.5%;  Pred. No. 4.6e-71;
  Matches  427;  Conservative    0;  Mismatches   24;  Indels    1;  Gaps    1;

Qy   1299 ctcacggggttggtggagtcccttatgttctctcaccagcctacttgggatgattgtcaa 1358
           ||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTCACGGCGTTAGTGGAGTCCCTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAG 60

Qy   1359 cagctgctgcagacactcttcacaaccgaggagcgagagagaattctgttagaggctaga 1418
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CAGCTGCTGCAGACACTCTTCACAACCGAGGAGCGAGAGAGAATTCTGTTAGAGGCTAG- 119

Qy    1419 aaaaatgttcctggggccgacgggcgacccacgcagttgcaaaatgagattgacatggga 1478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 AAAAATGTTCCTGGGGCCGACGGGCGACCCACGCAGTTGCAAAATGAGATTGACATGGGA 179

Qy    1479 tttcccttgactcgccccggttgggactacaacacggctgaaggtagggagagcttgaaa 1538
          |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||
Db     180 TTTCCCTTGACTCGCCCCGGATGGGACTACAACACGGCCGAAGGTAGGGAGAGCTTGAAA 239

Qy    1539 atctatcgccaggctctggtggcgggtctccggggcgcctcaagacggcccactaatttg 1598
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     240 ATCTATCGCCAGGCTCTGGTGGCGGGTCTCCGAGGCGCCTCAAGACGGCCCACTAATTTG 299

Qy    1599 gctaaggtaagagaggtgatgcagggaccgaacgaacctccctcggtatttcttgagagg 1658
          |||||||||||||  |||||||||||| ||| ||||| |||| || ||||||||||||
Db     300 GCTAAGGTAAGAGAAATGATGCAGGGACTGAATGAACCCCCCTCTGTTTTTCTTGAGAGG 359

Qy    1659 ctcatggaagccttcaggcggttcacccctttgatcctacctcagaggcccagaaagcc 1718
          ||| |||||||||||||||| |||||||||||||| |||||||||| | ||||||
Db     360 CTCTTGGAAGCCTTCAGGCGGTACACCCCTTTTGATCCCACCTCAGAGGCCTAAAAAGCC 419

Qy    1719 tcagtggccctggccttcattgggcagtcggc 1750
          ||||||||  ||||||| || || ||||| ||
Db     420 TCAGTGGCTTTGGCCTTTATAGGACAGTCAGC 451


RESULT  11
AZ730616
LOCUS       AZ730616      576 bp    DNA             GSS       25-JAN-2001
DEFINITION  RPCI-24-149D1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-149D1,
            DNA sequence.
ACCESSION   AZ730616
VERSION     AZ730616.1  GI:12491859
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-149D1.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 149  row: D  column: 1
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-149D1"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      161 a    139 c    121 g    155 t
ORIGIN


  Query Match             4.7%;  Score 385;  DB 13;  Length 576;
  Best Local Similarity   82.5%;  Pred. No. 1e-67;
  Matches  453;  Conservative    0;  Mismatches   95;  Indels    1;  Gaps    1;

Qy    5040 ggaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccaggtaagtcag 5099
          || ||||| |||||||| |||||     ||||| || || || |||||| ||||||||||
Db       1 GGCATACCCAAGGTAATCGGGTCCATTAATGGACCTGCCTTTGTTGCCAAGGTAAGTCAG 60

Qy    5100 ggactggccaagatattggggattgattggaaactgcattgtgcatacagaccccaaagc 5159
          ||||| |||||||||||||||||||||||| |||  | ||||||| |||||| | ||||| |
Db      61 GGACTAGCCAAGATATTGGGGATTGATTAGAAGTTACATTGTGTATACAGGCTCCAAAAC 120

Qy    5160 tcaggacaggtagagaggatgaatagaaccattaaagagacccttaccaaattgaccaca 5219
          |||||||||||||||||||||||||||||| |||||| |||||||||||||||||||| |  |
Db     121 TCAGGACAGGTAGAGAGGATGAATAAAACCATAAAAGAGACCCTTACCAAATTG-CTGCT 179

Qy    5220 gagactggcattaatgattggatggctctcctgccctttgtgctttttagggtgaggaac 5279
            ||||| || |||||||||||| |||||||| |||||||||| || ||||| || |||
Db     180 ATGACTGCCACTAATGATTGGATAGCTCTCCTACCCTTTGTGCTCTTCAGGGTTAGAAAC 239

Qy    5280 acccctggacagtttgggctgacccccctatgaattgctctacgggggaccccccccgttg 5339
          |||||||||||||| | |||| ||||||||||  ||||| |||||  || | || |||
Db     240 ACCCCTGGACAGTTTAGACTGAACCCCTATGAATCACTCTATGGGGGGTCCTCTCCATTG 299

Qy    5340 gcagaaattgcctttgcacatagtgctgatgtgctgctttcccagcctttgttctctagg 5399
          | | |||| |||| || ||||||||||||||||||||||||||||||||||||||||||||
Db     300 GTAAAAATAGCCTCTGTACATAGTGCTGATGTGCTGCTTTCCCAGCCTTTGTTCTCTAGT 359

Qy    5400 ctcaaggcgctcgagtgggtgaggcagcgagcgtggaagcagctccgggaggcctactca 5459
          ||||||| |||||| |||||||| |||| ||  | |||||||||||  | | |||||||||
Db     360 CTCAAGGTGCTCGAATGGGTGAGACAGCAAGTATAGAAGCAGCTCCAGAATGCCTACTCA 419

Qy    5460 ggaggagacttgcaagttccacatcgcttccaagttggagattcagtctatgttagacgc 5519
           ||| |||| || |||  |||||||  ||||||||||  ||||||||||||  ||||||||
Db     420 AGAGAAGACCTGGAAGCCCCACATCTCTTCCAAGTTATAGATTCAGTCTACATTAGACGC 479

Qy    5520 caccgtgcaggaaacctcgagactcggtggaagggaccttatctcgtacttttgaccaca 5579
          | ||   ||| |||| |  |||||| | |||| || |||||||| ||| | || |||||
Db     480 CTCCACACAGAAAACTTTAAGACTCAGCGGAAAGGCCCTTATCTTGTAATCTTTACCACC 539

Qy    5580 ccaacggct 5588
          | ||   ||
Db     540 CTAATTTCT 548


RESULT  12
BI182930/c
LOCUS       BI182930      564 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-br-h-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-br-h-08-0-UNL 3', mRNA sequence.
ACCESSION   BI182930
VERSION     BI182930.1  GI:14657339
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
```

```
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized porcine ovarian follicles library
            Seq primer: M13 -29
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .564
                      /organism="Sus scrofa"
                      /strain="University of Nebraska, Lincoln Swine Selection
                      Lines"
                      /db_xref="taxon:9823"
                      /clone="UNL-P-FN-br-h-08-0-UNL"
                      /clone_lib="UNL-P-FN"
                      /dev_stage="ADULT"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                      library is a normalized library representing porcine
                      ovarian follicles, ranging between 2.0 to 10.0 mm in
                      diameter, collected  during 7 days of the follicular phase
                      of the pig estrous cycle. This  library was derived from
                      the library UNL-P-F2. The tag is a string of  5-6
                      nucleotides present between the Not I site and the
                      oligo-dT track.  The library was constructed as described
                      by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                      , 1996.
                      TAG_LIB=UNL-P-FN
                      TAG_TISSUE=porcine ovarian follicles
                      TAG_SEQ=CACACT"
BASE COUNT      141 a    129 c    143 g    151 t
ORIGIN


  Query Match             4.7%;  Score 383.6;  DB 11;  Length 564;
  Best Local Similarity   84.6%;  Pred. No. 2e-67;
  Matches  477;  Conservative    0;  Mismatches    9;  Indels   78;  Gaps    1;

Qy   7720 cctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggcctgcta 7779
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    564 CCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTA 505

Qy   7780 tcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggattctgtaa 7839
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    504 TCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAA 445

Qy   7840 aactgactggcaccatagaagaattgattacacattgacagccctagtgacctatctcaa 7899
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    444 AACTGACTGGCACCATAGAAGAACTGATTACACATTGACAGCCCTAGTGACCTATCTCAA 385

Qy   7900 ctgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagctattttaa 7959
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    384 CTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGACCTCCAGAGCTATTTTAA 325

Qy   7960 aatgattggtccacggagcgcgggctctcgatattttaaaatgattggtccatggagcgc 8019
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    324 AATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCACGGAGCGC 265

Qy   8020 g------------------------------------------------------------ 8020
           |
Db    264 GGGCTCTCGATATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAAT 205

Qy   8021 -------------------ggctctcgatattttaaaatgattggtttgtgacgcacagg 8061
                              ||||||||||||||||||||||||||||||||||||||||
Db    204 GATTGGTCCACGAAGCGCCGGCTCTCGATATTTTAAAATGATTGGTTTGTGACGCACAGG 145

Qy   8062 ctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgcagtcctctac 8121
           |||||||||||||||||||||||||||||||| ||||||| ||  ||||||||||||||
```

```
Db    144 CTTTGTTGTGAACCCCATAAAAGCTGTCCCGACTCCGCACACGCCGCCGCAGTCCTCTAC 85

Qy   8122 ccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtt 8181
           ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db     84 CCCTGCGTGGCCTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTT 25

Qy   8182 tgcatcaaaaaaaaaaaaaaaaaa 8205
           |||||||||||||||||||||||||
Db     24 TGCATCAAAAAAAAAAAAAAAAAA 1


RESULT  13
BH056750
LOCUS       BH056750      864 bp    DNA             GSS       18-JUL-2001
DEFINITION  RPCI-24-94C16.TVB RPCI-24 Mus musculus genomic clone RPCI-24-94C16,
            DNA sequence.
ACCESSION   BH056750
VERSION     BH056750.1  GI:14864050
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 864)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-94C16.TJ RPCI-24-94C16.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 94  row: C  column: 16
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .864
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="RPCI-24-94C16"
                      /clone_lib="RPCI-24"
                      /sex="Male"
                      /cell_type="Spleen/Brain"
                      /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                      RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                      library was  cloned in the pTARBAC1 cloning vector at the
                      BamH1 sites using MboI  partially digested male C57BL/6J
                      DNA."
BASE COUNT      237 a    221 c    209 g    197 t
ORIGIN


  Query Match             4.6%;  Score 380.8;  DB 13;  Length 864;
  Best Local Similarity   66.5%;  Pred. No. 7e-67;
  Matches  577;  Conservative    0;  Mismatches  287;  Indels    4;  Gaps    2;

Qy   3414 gtgagagagtttttggggacagctggattttgcagactgtggatcccgggggtttgcgacc 3473
           ||  | ||||||  |||| || || || |||||  |||||  || ||  || ||||   | |
Db      1 GTACGTGAGTTTCTGGGAACGGCAGGCTTTTGTAGACTCTGAATACCAGGGTCCACAATG 60

Qy   3474 ttagcagccccactctacccgctaaccaaagaaaaaggggattctcctgggctcctgag 3533
           |||||  ||  ||  ||  |||||  ||  ||  ||||||||||     |||  |||||  |    ||
```

```
Db      61 TTAGCGGCACCTCTGTACCCACTCACTAAAGAAAAAGTCCCGTTCACCTGGACAGAAGAA 120

Qy    3534 caccagaaggcatttgatgctatcaaaaaggccctgctgagcgcacctgctctggccctc 3593
           ||||| || |  ||| |    || ||    || |||||    || |  ||| |||||||
Db     121 CACCAAAAAGACTTTAACAGCATAAAGGCTGCTCTGCTTGCAGCCCAGGCTTTGGCCCTG 180

Qy    3594 cctgacgtaactaaaccctttaccctttatgtggatgagcgtaagggagtagcccgagga 3653
           ||  |  | ||||| || ||  ||| | || || |||||||     ||||||||||  ||
Db     181 CCATATTTGACTAAGCCTTTCACCATATACGTTGATGAGCCAGCTGGAGTAGCCCGCAGA 240

Qy    3654 gttttaacccaaaccctaggaccatggaggagacctgttgcctacctgtcaaagaagctt 3713
           ||  | || || || || || || |||| ||| |  || |||||||||| || ||  |
Db     241 GTCCTGACTCAGACTCTGGGGCCGTGGAAGAGGCTAGTGGCCTACCTGTCTAAAAAATTG 300

Qy    3714 gatcctgtagccagtggttggcccgtatgtctgaaggctatcgcagctgtggccatactg 3773
           || || || |||||||  ||||||   || || || || ||||| || || ||| ||||
Db     301 GACCCAGTTGCCAGTGAATGGCCCTCCTGCCTAAAAGCCATCGCTGCAGTAGCCCTACTT 360

Qy    3774 gtcaaggacgctgacaaattgactttgggacagaatataactgtaatagccccccatgca 3833
           || || ||  ||| |||  | ||| | || ||| || | | ||||||  | ||||||||
Db     361 GTTAAAGATACTGGCAAGCTTACTCTAGGTCAGCATGTGAGTGTAATCACTCCCCATGCG 420

Qy    3834 ttggagaacatcgttcggcagcccccagaccgatggatgaccaacgcccgcatgacccac 3893
           ||  |||  || ||  | ||||||| |||| |||||||| || ||||| |||||||||
Db     421 TTAAAGAGTATTGTATGACAGCCCCCTGACCATTGGATGACAAATGCCCGGATGACCCAC 480

Qy    3894 tatcaaagcctgcttctcacagagagggtcactttcgctccaccagccgctctcaaccct 3953
           || || |||||| | || |  ||  | || ||||| || | ||| |||   | ||| |||
Db     481 TACCAGAGCCTGTTGCTAAATGACCGAGTGACTTTTGCCCTACCTGCCATCCACAATCCT 540

Qy    3954 gccactcttctgcctgaagagactgatgaaccagtgactcatgattgccatcaactattg 4013
            |||| ||||| ||  |  |  |||     |  |  ||  | ||  | | |  | ||
Db     541 ACCACCCTTCTCCCCAAGACGGATGACTCTACCTGGGTACACCAATGTGCTGATATCCTG 600

Qy    4014 attgaggagactggggtccgcaaggaccttacagacataccgctgactggagaagtgcta 4073
             ||| || ||||||     | ||||| || || |||  | |   | |   ||  ||||
Db     601 GATGAAGAAACTGGATCGAGAAAGGATCTGACTGACCAAACTTGGCC---AGGTGTGCCT 657

Qy    4074 acctggttcactgacggaagcagctatgtggtggaaggtaagaggatggctggggcggca 4133
            | ||| ||| || || ||||||| |||||||| || || |  |  || || | ||||
Db     658 GACAGGTACACAGATGGTAGCAGCTTTGTGGTGGAGGGGAAAATAAGAGCAGGAGAGGCA 717

Qy    4134 gtggtggacgggacccgcacgatctgggccagcagcctgccggaaggaacttcagcgcaa 4193
           ||||| ||| | |  |     || ||||| ||    || || ||||||||||| || ||
Db     718 GTGGTTGACAGAAAACATGTAATTTGGGCAAGTTAACTTCCAGAAGGAACTTCCGCTCAG 777

Qy    4194 aaggctgagctcatggccctcacgcaagctttgcggctggccgaagggaaatccataaac 4253
           |||||||||||| || | || ||  ||||| ||| ||||| | || || |  | || ||
Db     778 AAGGCTGAGCTCCTGACGCTAACTAAAGCTCTGCCGCTGG-CAAAAGGTAGGCAATCAAT 836

Qy    4254 atttatacggacagcaggtatgcctttg 4281
           ||||| || ||||||   ||||| ||||
Db     837 ATTTACACTGACAGCCACTATGCTTTTG 864
```

```
RESULT  14
AZ318021
LOCUS       AZ318021      729 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0036O04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0036O04 R, DNA sequence.
ACCESSION   AZ318021
VERSION     AZ318021.1  GI:10367383
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0036  row: O  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 729.
FEATURES            Location/Qualifiers
     source         1. .729
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UUGC1M0036O04"
                    /clone_lib="Mouse 10kb plasmid UUGC1M library"
                    /sex="Male"
                    /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                    /note="Vector: PWD42nv; Purified genomic DNA from M.
                    musculus C57BL/6J (male) was obtained  from the Jackson
                    Laboratory Mouse DNA Resource
                    (http://www.jax.org/resources/documents/dnares/). The DNA
                    was hydrodynamically sheared by repeated passage through a
                    0.005 inch orifice at constant velocity. The sheared DNA
                    was blunt end-repaired with T4 DNA polymerase and T4
                    polynucleotide kinase. Adaptor oligonucleotides were
                    ligated to the blunt ends in high  molar excess. The
                    adaptored DNA was purified and size-selected for a 9.5 to
                    10.5 kb range using preparative agarose gel
                    electrophoresis. Vector DNA was prepared from a derivative
                    of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                    inducible derivative of plasmid R1. The vector was ligated
                    with adaptors complementary to the insert adaptors and
                    purified. The sheared, adaptored mouse DNA was annealed to
                    adaptored vector DNA, and transformed into
                    chemically-competent E. coli XL10-Gold (Stratagene) cells
                    and selected for ampicillin resistance."
BASE COUNT      196 a    193 c    182 g    158 t
ORIGIN
```

```
  Query Match             4.6%;  Score 380.4;  DB 13;  Length 729;
  Best Local Similarity   73.4%;  Pred. No. 8.5e-67;
  Matches  486;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy    1215 cccctccagtattggccctttttcttctgcagatctctataattggaaaactaaccatccc 1274
           ||  | || || ||||| ||||| ||| | ||| | ||||||||||||||||||||| ||
Db      65 CCTTTACAATACTGGCCTTTTTCCTCTTCTGATTTATATAATTGGAAAACTAACCACCCT 124

Qy    1275 cctttctcggaggatccccaacgcctcacggggttggtggagtcccttatgttctctcac 1334
           || ||||| ||| | |||    | || || ||| |  | ||||| |||||||||| ||
Db     125 CCCTTCTCAGAGAACCCCTCTGGGCTTACTGGGCTCCTTGAGTCACTTATGTTCTCCCAT 184

Qy    1335 cagcctacttgggatgattgtcaacagctgctgcagacactcttcacaaccgaggagcga 1394
           || || |||||||||||||||||| |||||  |||||   || ||||| || || ||  ||
Db     185 CAACCCACTTGGGATGATTGTCAGCAGCTTTTGCAGGTTCTTTTCACCACAGAAGAAAGA 244

Qy    1395 gagagaattctgttagaggctagaaaaaatgttcctggggccgacgggcgacccacgcag 1454
           || ||||| ||| | ||||| |||||||||||||  || |  ||||    ||||||
Db     245 GAAAGAATCCTGATGGAGGCGAGAAAAAATGTTCTAGGAGAGGACGACACACCCACTGCC 304

Qy    1455 ttgcaaaatgagattgacatgggatttcccttgactcgccccggttgggactacaacacg 1514
            | |  ||     | |||  ||  || |||||||  ||||||   ||||||||||||||
Db     305 CTCCCTAACCTCGTGGACGAGGCTTTCCCCTTGAACCGCCCCAACTGGGACTACAACACC 364

Qy    1515 gctgaaggtagggagagcttgaaaatctatcgccaggctctggtggcgggtctccggggc 1574
           || ||||||||||   || |     ||||||||| | |||| ||||| |||||| | ||
```

```
Db     365 GCGGAAGGTAGGGGACGCCTCCTTGTCTATCGCCGGACTCTAGTGGCAGGTCTCAGAGGA 424

Qy    1575 gcctcaagacggcccactaatttggctaaggtaagagaggtgatgcagggaccgaacgaa 1634
           ||| | |||||||||||| ||||||||||||||||||||||||  ||||||| | ||  |||
Db     425 GCCGCTAGACGGCCCACCAATTTGGCTAAGGTAAGAGAGGTCTTGCAGGGGCAGACTGAA 484

Qy    1635 cctccctcggtatttcttgagaggctcatggaagccttcaggcggttcacccctttttgat 1694
           || ||||| || || |||||| | |||||||| || |  | | ||| ||||||||||||
Db     485 CCACCCTCAGTCTTCCTTGAGCGTCTCATGGAGGCATATAAGAGGTACACCCCTTTTGAC 544

Qy    1695 cctacctcagaggcccagaaagcctcagtggccctggccttcattgggcagtcggctctg 1754
           ||  | |||||||  || ||||||  | || ||| ||||||||||||| ||||| ||||
Db     545 CCCTCGTCAGAGGGGCAAAAAGCCGCTGTAGCCATGGCCTTCATTGGTCAGTCCGCTCCC 604

Qy    1755 gatatcaggaagaaacttcagagactggaagggttacaggaggctgagttacgtgatcta 1814
           || || | ||| || || || || ||||| ||| | || ||   |  | | |  ||| ||
Db     605 GACATTAAGAAAAAGCTGCAAAGGCTGGAGGGGCTCCAAGATCATACGCTCCAAGATTTA 664

Qy    1815 gtgagagaggcagagaaggtgtattacagaagggagacagaagaggagaaggaacagaga 1874
           || | ||| |||||||| || ||| | |  ||||| |||||||| |||| | |  |||||
Db     665 GTAAAAGAAGCAGAGAAAGTCTATCATAAGAGGGAAACAGAAGAAGAGAGGCAGGAGAGA 724

Qy    1875 aa 1876
             |
Db     725 GA 726
```

```
RESULT  15
AZ329067/c
LOCUS       AZ329067      673 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0053G10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0053G10 F, DNA sequence.
ACCESSION   AZ329067
VERSION     AZ329067.1  GI:10389414
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0053  row: G  column: 10
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 673.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0053G10"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      165 a    173 c    149 g    186 t
ORIGIN
```

```
  Query Match             4.5%;  Score 373.2;  DB 13;  Length 673;
  Best Local Similarity   78.1%;  Pred. No. 2.5e-65;
  Matches  520;  Conservative    0;  Mismatches  123;  Indels   23;  Gaps    5;

Qy    4851 ctaaggggaagccacccaggcgctcactgggaagtggacttcactgaggtaaagccggct 4910
           || ||||   ||   || |  ||  ||||||| || || || || ||    ||||| |||
Db     672 CTCAGGGATGGCTGGCCTGAAGCCTACTGGGAGGTAGAGTTTACAGAATCTAAGCCAGCT 613

Qy    4911 aaatacggaaacaaatatctattggtttttgtagacacctttttcaggatgggtagaggct 4970
           | |||||| ||||| ||  |||| || ||| |||| ||||||||| | ||| |||| |||
Db     612 AGATACGGCAACAAGTACTTATTAGTGTTTATAGATACCTTTTCAAGGTGGATAGAAGCT 553

Qy    4971 tatcctactaagaaagagacttcaaccgtggtggctaagaaaatactggaggaaatttttt 5030
           | |||||| | ||  |||||           ||| || |  || || || ||||| ||
Db     552 TTTCCTACCAGGACTGAGAC------------GGCCAATATGATCCTAGACGAAATCTTC 505

Qy    5031 ccaagatttggaatacctaaggtaatagggtcagacaatggtccagctttcgttgcccag 5090
           || || ||||  ||||| ||||||||  ||||  | ||||| || || || |||||||||
Db     504 CCGAGGTTTGTGATACCCAAGGTAATCAGGTCCAATAATGGACCTGCCTTTGTTGCCCAG 445

Qy    5091 gtaagtcagggactggccaagatattggggattgattggaaactgcattgtgcatacaga 5150
           ||||||||||||||||||||||||| | ||||| || |||  | ||||||||||||||
Db     444 GTAAGTCAGGGACTGGCCAAGATATTAGAGATTGGTTAGAAGTTACATTGTGCATACAGG 385

Qy    5151 ccccaaagctcaggacaggtagagaggatgaatagaaccattaaagagacccttaccaaa 5210
           ||||||||||||||||||||||||| | |||||||||||| ||||||||||||||||||
Db     384 CCCCAAAGCTCAGGACAGGTAGAGAGAACGAATAGAACCATT-AAGAGACCCTTACCAAA 326

Qy    5211 ttgaccacagagactggcattaatgattggatggctctcctgccctttgtgctttttagg 5270
           ||| |  | |||||| |   |||||||  ||| |||||||| |||||||||| || |||
Db     325 TTGGCAGCTGAGACTAGAGCTAATGATAAGATAGCTCTCCTACCCTTTGTGCTCTTCAGG 266

Qy    5271 gtgaggaacacccctggacagtttgggctgacccccctatgaattgctctac----ggggg 5326
           || || |||| ||||| ||||||||| |||||||||||| |||| ||||||    |||||
Db     265 GTTAGAAACATCCCTGAACAGTTTGGACTGACCCCCTATAAATTACTCTACGGGGGGGGG 206

Qy    5327 acccccccccgttggcagaaattgcctttgcacatagtgctgatgtgctgctttcccagcc 5386
                 | ||  ||| | |||| |||| || ||||||||||||  ||||||||||||||||
Db     205 GGGGGCTCCACTGGTAAAAATAGCCTCTGTACATAGTGCTGACATGCTGCTTTCCCAGCC 146

Qy    5387 tttgttctctaggctcaaggcgctcgagtgggtgaggcagcga----gcgtggaagcagc 5442
           |||||||||||||||||| |||||||||||||||| ||| ||    || ||||||||||
Db     145 TTTGTTCTCTAGGCTCAAGGTGCTCGAGTGGGTGAGACAGTGAGTGGGCATGGAAGCAGC 86

Qy    5443 tccgggaggcctactcaggag--gagacttgcaagttccacatcgcttccaagttggaga 5500
           ||| | |||||||||| |||  ||||| ||||||  ||||||| ||| ||||||| ||||
Db      85 TCCAGAAGGCCTACTCAAGAGAAGAGACCTGCAAGCCCCACATCTCTTACAAGTTAGAGA 26

Qy    5501 ttcagt 5506
           ||| ||
Db      25 TTCGGT 20
```

Search completed: February 23, 2002, 23:37:03
Job time: 9424 sec

OM of: US-09-171-553B-4 to: EST:* out_format : pfs

Date: Feb 24, 2002 8:31 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09171553/runat_22022002_122825_11137/app_query.fasta_1.3312
-DB=EST -QFMT=fastap -SUFFIX=feb22std.rst -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09171553_@CGN1_1_9008 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-171-553B-4
Query length: 524
Database: EST:*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 3515.700000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_htc:AK011989	+	1559.50	1582.89	5.3e-79	2096	! AK011989 Mus musculus 10 days
gb_htc:AK014027	+	1523.00	1545.85	6.1e-77	2085	! AK014027 Mus musculus 13 days
gb_htc:AK004927	+	1223.50	1238.89	7.7e-60	2935	! AK004927 Mus musculus adult ma
gb_gss:BH063533	+	991.00	1012.63	3.1e-47	801	! BH063533 RPCI-24-370H8.TJ RPCI-
gb_gss:AZ318021	+	946.00	967.63	9.8e-45	729	! AZ318021 1M0036O04R Mouse 10kb
gb_gss:BH121259	+	916.00	937.70	4.6e-43	679	! BH121259 RPCI-24-374G11.TV RPCI
gb_gss:AZ967275	+	861.00	882.19	5.6e-40	646	! AZ967275 2M0238G10F Mouse 10kb
gb_gss:BH124013	+	846.00	866.98	4.0e-39	644	! BH124013 RPCI-24-230P16.TV RPCI
gb_gss:AZ420279	-	845.50	866.80	4.1e-39	617	! AZ420279 1M0198K05F Mouse 10kb
gb_gss:AZ986580	+	828.50	848.81	4.1e-38	677	! AZ986580 2M0268P18R Mouse 10kb
gb_gss:AZ862103	-	827.00	847.28	5.0e-38	677	! AZ862103 2M0169D03R Mouse 10kb
gb_est2:BG089942	-	810.50	830.89	4.1e-37	645	! BG089942 mab88d06.x1 NCI_CGAP_S
gb_est1:BE283560	+	809.00	829.95	4.6e-37	598	! BE283560 601102760F1 NCI_CGAP_I
gb_gss:AZ650638	+	800.00	820.77	1.5e-36	601	! AZ650638 1M0520L20R Mouse 10kb
gb_gss:AZ708792	+	797.50	815.83	2.8e-36	820	! AZ708792 RPCI-23-238P4.TJ RPCI-
gb_est1:BE283558	+	796.00	816.91	2.4e-36	585	! BE283558 601102758F1 NCI_CGAP_I
gb_est2:BF143092	+	793.00	810.61	5.5e-36	892	! BF143092 601788777F1 NCI_CGAP_I
gb_gss:AZ777287	+	787.50	806.72	9.0e-36	716	! AZ777287 2M0011P01R Mouse 10kb
gb_gss:AZ866922	+	783.00	803.62	1.3e-35	591	! AZ866922 2M0177M06R Mouse 10kb
gb_gss:AZ799523	+	782.00	802.63	1.5e-35	589	! AZ799523 2M0057L01F Mouse 10kb
gb_gss:AZ348675	-	782.00	802.58	1.5e-35	593	! AZ348675 1M0085P17F Mouse 10kb
gb_gss:AZ970796	-	776.50	795.95	3.6e-35	679	! AZ970796 2M0244L04F Mouse 10kb
gb_est2:BF450334	-	776.00	796.21	3.5e-35	614	! BF450334 uz77c10.x1 NCI_CGAP_Lu
gb_gss:BH112876	+	773.00	793.98	4.6e-35	552	! BH112876 RPCI-24-368D12.TJ RPCI
gb_gss:BH086028	+	764.50	784.19	1.6e-34	642	! BH086028 RPCI-24-338O16.TV RPCI
gb_est2:BF441468	+	762.00	784.11	1.6e-34	466	! BF441468 257862 MARC 2PIG Sus s
gb_est2:BF441465	+	762.00	784.09	1.6e-34	467	! BF441465 257856 MARC 2PIG Sus s
gb_est1:AW763854	-	745.00	765.05	1.9e-33	588	! AW763854 ur66f07.x1 NCI_CGAP_Ma
gb_gss:AZ859749	-	744.00	764.37	2.1e-33	563	! AZ859749 2M0165A03R Mouse 10kb
gb_est2:BF450406	-	728.00	747.59	1.8e-32	602	! BF450406 uz78b07.x1 NCI_CGAP_Lu
gb_gss:BH061935	+	727.00	747.33	1.8e-32	546	! BH061935 RPCI-24-370D19.TJ RPCI
gb_gss:AZ385280	+	727.00	746.39	2.1e-32	617	! AZ385280 1M0143L11R Mouse 10kb
gb_gss:AZ325348	+	727.00	746.21	2.1e-32	631	! AZ325348 1M0047N11R Mouse 10kb
gb_gss:AZ980174	-	723.50	742.17	3.6e-32	672	! AZ980174 2M0257D10F Mouse 10kb
gb_est1:AW215497	-	718.50	736.74	7.1e-32	704	! AW215497 up04h01.x1 NCI_CGAP_Lu
gb_est2:BF441466	+	718.00	739.64	4.9e-32	452	! BF441466 257857 MARC 2PIG Sus s
gb_est2:BF659611	-	715.50	734.28	9.8e-32	652	! BF659611 uz88f01.x1 NCI_CGAP_Lu
gb_est2:BF148613	-	714.00	733.41	1.1e-31	599	! BF148613 uy32b01.x1 NCI_CGAP_Lu
gb_est2:BG921696	+	713.00	730.07	1.7e-31	810	! BG921696 602825515F1 NCI_CGAP_M
gb_gss:AZ657594	+	702.00	720.97	5.4e-31	618	! AZ657594 1M0533F21R Mouse 10kb
gb_gss:AZ959168	-	700.00	719.30	6.7e-31	590	! AZ959168 2M0226C14R Mouse
gb_gss:AZ937454	-	697.50	716.23	9.9e-31	632	! AZ937454 2M0195G09R Mouse
gb_est2:BI081838	+	695.50	711.79	1.8e-30	864	! BI081838 602879750F1 NCI_C
gb_gss:AZ821842	-	690.50	709.15	2.5e-30	629	! AZ821842 2M0094H15R Mouse
gb_htc:AK007448	+	690.50	707.13	3.2e-30	818	! AK007448 Mus musculus 10 d

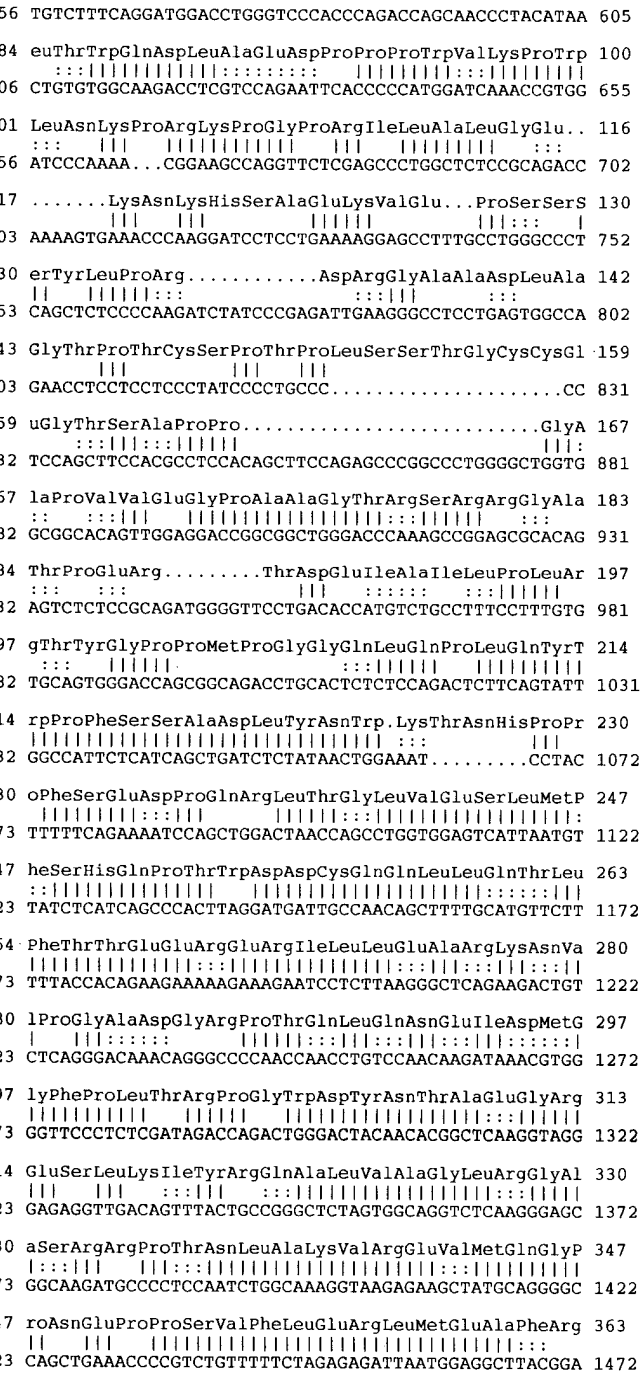
seq_name: gb_htc:AK011989

seq_documentation_block:

LOCUS AK011989 2096 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus 10 days embryo cDNA, RIKEN full-length enriched library, clone:2610305J24, full insert sequence.
ACCESSION AK011989
VERSION AK011989.1 GI:12848454
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2610305J24.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2096)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 2096)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 2096)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (bases 1 to 2096)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 2096)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

```
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .2096
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1905349"
                     /db_xref="MGD:MGI:1919770"
                     /clone="2610305J24"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days embryo"
     CDS             688. .>2096
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB27961.1"
                     /db_xref="GI:12848455"
                     /translation="MGQTVSTPLSLTKDHWTDVRARGQNLSVKVKKKPWMTFCSSEWP
                     VFGVGWPAEGTFYLPTIRAVKAIVFQEGPGSHPDQQPYIMVWEDLARYPPPWVRPFLP
                     PLRPGTKILAIQENGEKEKPKPPLGRDDDRSTPVTKPPKIYSEIEEPPEWPDTPQPPP
                     YAPQPQPSAPSGPLPQAPAGGGPSTGTRSRRGVTPEGPADSTVALPLRAIGAPPADP
                     NSLQPLQYWPFSSSDLYNWKAHHPPFSENPAGLTGLVESLMYSHQPTWDDCQQLLQTL
                     FTTEERERILLEARKNVRDEAGRPVQTPAEIDEGFPLTRPRWDYNTASGRERLSNYRR
                     VLVAGLRGAARQPTNLAKVREVMQGATEPPSVFLERLMEAYRRYTPFDPTSEGQRASV
                     IMAFIGQSAPDIRKKLQRIEGLQDYTIRDVVREAEKVYHRRETEDKKLEREKREKREE
                     EDRRDRRQEKVLTRILAAVG"
BASE COUNT      500 a    527 c    577 g    492 t
ORIGIN


alignment_scores:
          Quality: 1559.50              Length:    473
            Ratio:  4.051                Gaps:      7
 Percent Similarity:  81.395   Percent Identity: 65.328

alignment_block:
 US-09-171-553B-4 x AK011989   ..

 Align seg 1/1  to: AK011989  from: 1  to: 2096

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         |||||||||||||||:::|||||||||||||||||||   ||||||||||
     688 ATGGGACAGACCGTGTCTACTCCTTTATCTTTGACTAAGGACCATTGGAC 737

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
         |:::||||||:::|||:::::|||||||||||:::|||||||||   |
     738 GGACGTTAGGGCTAGAGGACAAAATTTGTCAGTAAAAGTGAAGAAAAAGC 787

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
         |||||:::||||||||||:::|||||||||||:::|||   |||||||||
     788 CATGGATGACTTTCTGTTCCTCAGAATGGCCTGTTTTTGGAGTAGGTTGG 837

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||:::|||||||||||||             |||   ||||||||||||||
     838 CCAGCAGAAGGAACTTTTTACTTACCCACCATAAGGGCTGTGAAGGCCAT 887

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
         |:::||||||   |||||||||||||||||||||||:::|||||||||:
     888 TGTTTTTCAGGAAGGACCAGGGTCGCATCCAGACCAACAACCGTACATCA 937

      84 euThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
         :::::|||:::||||||||::   |||||||||||||||:::|||:::
     938 TGGTATGGGAGGACTTGGCACGCTACCCACCCCCGTGGGTTCGCCCATTC 987

     101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeu........ 114
         |||      |||   :::|||||||   :::||||||||:::
     988 CTC...CCGCCTCTCCGCCCTGGCACCAAGATTCTAGCCATCCAAGAAAA 1034

     115 .GlyGluLysAsnLys.............................HisS 121
          |||||||||:::|||                             :::|
    1035 TGGTGAGAAAGAGAAACCGAAACCACCGCTCGGGAGAGATGATGATCGCA 1084

     121 erAlaGluLysValGluProSerSerSerTyr......LeuProArgAsp 135

         ||:::      ::::::|||   :::   |||      :::   :::
    1085 GCACACCAGTGACGAAACCCCCCAAGATCTATTCAGAGATTGAAGAACCC 1134

     136 ArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThrProLe 152
                 |||           |||      :::|||   |||
    1135 CCTGAGTGGCCCGATACCCCTCAACCCCCACCTTATGCTCCCCAGCCCCA 1184

     152 uSerSerThrGlyCysCysGluGlyThrSerAlaProProGlyAlaProV 169
          |||:::                 :::|||||||   |||:::
    1185 ACCTTCAGCTCCCTCGGGACCCCTGCCTCAGGCCCCGGCCGGAGGA.... 1230

     169 alValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAlaThrPro 185
               ||||||:::::|||||||||||||||||||||||:::||||||
    1231 .....GGGGGTCCCTCCACAGGAACAAGGAGCCGGCGAGGAGTCACCCCT 1275

     186 Glu...ArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGly.. 200
          |||      :::|||:::   :::   ||||||||||||:::   |||
    1276 GAGGGGCCTGCGGATTCAACTGTGGCGCTCCCCCTCAGGGCTATTGGGGC 1325

     201 .ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheS 217
          ||||||          ::::::|||||||||||||||||||||||||||
    1326 TCCCCCTGCCGATCCAAATAGTCTACAGCCCCTACAGTATTGGCCTTTTT 1375

     217 erSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
         |||||:::|||||||||||||||||||:::::|||||||||||||||||||
    1376 CCTCTTCTGACCTTTACAACTGGAAAGCTCATCACCCCCCTTTTTCAGAA 1425

     234 AspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGl 250
         :::|||      |||||||||||||||||||||||||||||:::|||||||||
    1426 AACCCTGCAGGACTCACTGGGTTGGTTGAATCATTAATGTATTCACACCA 1475

     250 nProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrThrG 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1476 GCCGACCTGGGATGACTGCCAGCAGCTTCTGCAGACTCTATTCACAACCG 1525

     267 luGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAla 283
         ||||||||||||||||||||||||||||||||||||||||
    1526 AGGAGAGAGAGAGGATTCTCCTCGAGGCTCGGAAAAATGTCCGAGACGAG 1575

     284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLe 300
         |||||||||:::|||          |||||||||   ||||||||||
    1576 GCTGGGCGCCCTGTCCAAACTCCAGCTGAGATAGATGAAGGATTTCCGCT 1625

     300 uThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuL 317
         ||||||||||   ||||||||||||||||||:::|||||||||   |||:
    1626 AACCCGGCCCCGATGGGATTATAATACGGCATCAGGTAGGGAACGACTGT 1675

     317 ysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
         ::   ||||||:::::|||||||||||||||||||||||||||:::|||:::
    1676 CCAATTATCGCCGGGTCCTAGTGGCGGGTCTCAGAGGTGCTGCCCGGCAG 1725

     334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluPr 350
         ||||||||||||||||||||||||||||||||||||||||   :::||||||
    1726 CCCACGAATCTGGCCAAGGTAAGAGAGGTTATGCAGGGAGCGACTGAGCC 1775

     350 oProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrP 367
         |||||||||||||||||||||||||||||||||:::||||||:::||||
    1776 CCCCTCAGTCTTCCTTGAAAGGCTCATGGAGGCTTATAGGAGATATACCC 1825

     367 roPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPhe 383
         ||||||||||||||||||||:::|||:::|||||||||   :::|||||||
    1826 CATTCGACCCCACGTCTGAGGGTCAAAGGGCCTCAGTAATTATGGCCTTC 1875

     384 IleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGl 400
         ||||||||||||||   ||||||||||||||||||||||||:::||||||
    1876 ATTGGCCAGTCAGCTCCTGACATTAGGAAGAAGTTACAGCGAATTGAGGG 1925

     400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValT 417
         ||||||||:::      :::|||||||:::|||||||||||||||||||
    1926 CTTGCAGGATTACACCATAAGGGATGTAGTTAGAGAGGCAGAGAAAGTGT 1975

     417 yrTyrArgArgGluThrGluGluGluLysGluGlnArgLysGluLysGlu 433
         ||:::|||||||||||||||||:::::|||   :::|||:::::::::|||
```

```
    1976 ATCATAGGAGAGAAACAGAAGATAAAAAGTTAGAGAGAGAGAAAAGAGAG 2025

     434 ArgGluGluArgGluGluArgArgAspArgArgGlnGluLysAsnLeuTh 450
         ::::::|||:::|||:::||||||||||||||||||||||   |||||
    2026 AAAAGAGAAGAGGAGGATAGGAGAGACAGGAGGCAAGAAAAGGTTTTGAC 2075

     450 rLysIleLeuAlaAlaVal 456
         |:::|||||||||||||||
    2076 TAGGATCCTGGCCGCAGTA 2094

seq_name: gb_htc:AK014027

seq_documentation_block:
LOCUS       AK014027     2085 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus 13 days embryo head cDNA, RIKEN full-length enriched
            library, clone:3110009E22, full insert sequence.
ACCESSION   AK014027
VERSION     AK014027.1  GI:12851651
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 13 days embryo head cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:3110009E22.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2085)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 2085)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 2085)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 2085)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2085)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 50.0. Second strand cDNA was prepared with the primer adapter
            of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .2085
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1906029"
                     /db_xref="MGD:MGI:1920342"
                     /clone="3110009E22"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="13 days embryo"
     CDS             356. .1147
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB29122.1"
                     /db_xref="GI:12851652"
                     /translation="MGQTVTTPLSLTLDHWSEIRSRAHNLSVEIKKGPWRTFCASEWP
                     TFDVGWPPEGTFDLTVIFEVKAIVFQDGPGSHPDQQPYITVWQDLVQNSPPWIKPWIP
                     KRKPGSRALALRRPKVKPKDPPEKEPLPGPSALPKIYPEIEGPPEWPEPPPPYPLPPP
                     ASTPPQLPEPGPGAGGGTVGGPAAGTQSRSAQSLSADGVPDTMSAFPLCAVGPAADLH
                     SLQTLQYWPFSSADLYNWKSYFFRKSSWTNQPGGVINVISSAHLG"
BASE COUNT      544 a    504 c    573 g    464 t
ORIGIN


alignment_scores:
            Quality: 1523.00            Length:    555
              Ratio:  3.644              Gaps:     14
 Percent Similarity:  75.315   Percent Identity: 58.378

alignment_block:
 US-09-171-553B-4 x AK014027   ..

 Align seg 1/1  to: AK014027  from: 1  to: 2085

       1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||::
     356 ATGGGACAGACTGTGACGACCCCCCTAAGTTTGACCTTAGATCACTGGTC 405

      17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
         :|||:::|||||||||||||||||||||||||||::::::|||||||||
     406 TGAAATTAGGTCGAGGGCCCACAATCTTTCAGTGGAGATCAAGAAGGGAC 455

      34 roTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
         |||||:::|||||||||||||||||||||||||||||||||||||||||
     456 CTTGGCGGACTTTTTGTGCTTCTGAGTGGCCAACATTTGATGTGGGCTGG 505

      51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIl 67
         |||   ||||||||||||:::      :::|||:::   ||||||||||
     506 CCACCAGAGGGGACTTTTGACCTGACTGTGATCTTTGAAGTTAAAGCCAT 555

      67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
         |:::||||||   ||||||||||||||||||||||:::|||||||||
```



seq_name: gb_htc:AK004927

seq_documentation_block:

```
LOCUS       AK004927     2935 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male liver cDNA, RIKEN full-length enriched
            library, clone:1300007C21, full insert sequence.
ACCESSION   AK004927
VERSION     AK004927.1  GI:12836484
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:1300007C21.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2935)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 2935)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
```

```
   PUBMED   11042159
REFERENCE   3  (bases 1 to 2935)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 2935)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2935)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .2935
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1897042"
                     /db_xref="MGD:MGI:1914777"
                     /clone="1300007C21"
                     /sex="male"
                     /tissue_type="liver"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
BASE COUNT      669 a    860 c    773 g    633 t
ORIGIN


alignment_scores:
           Quality: 1223.50            Length:    563
             Ratio:  3.245               Gaps:     17
 Percent Similarity:  66.963   Percent Identity: 48.135
```

```
alignment_block:
 US-09-171-553B-4 x AK004927   ..

 Align seg 1/1  to: AK004927  from: 1  to: 2935

      1 MetGlyGlnThrValThrThr.ProLeuSerLeuThrLeuAspHisTrpT 17
        |||||||||||||||||||| ||||||||||||||||||||:::|||||| 
    639 ATGGGACAGACCGTAACTACCCCCCCTGAGTCTAACCTTGCAGCACTGGG 688

     17 hrGluValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGly 33
          :::|||:::      |||   |||   ||||||:::|||:::|||  
    689 GAGATGTCCAGCGCATTGCATCCAATCAGTCTGTGGATGTCAGGAAGAGG 738

     34 ProTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTr 50
          |||   |||||||||:::::|||||||||||||||:::|||||||| 
    739 CGCTGGGTTACTTTCTGTTCCGCCGAATGGCCAACTTTCAATGTGGGATG 788

     50 pProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaI 67
        ||||:::::|||:::|||||      ||||||      |||||||:::  
    789 GCCTCAGGATGGTGCTTTTAATTTGGGTATTATCTCTCAGGTCAAGTCTA 838

     67 leIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 83
          :::|||      ||||||   :::||||||||||||   |||||||||  
    839 GAGTGTTTTGTCCTGGTCCCCACGGACACCCGGATCAGGTCCCATATATC 888

     84 LeuThrTrpGlnAspLeuAla.GluAspProPro..............  94
        :::||||||:::   |||::: :::   ||||||               
    889 GTCACCTGGGAGGCACTCTCCAAAACCCCCTCCTTTACCGACAGCTCCCG 938

     94 ..................................................  94

    939 TCCTCCCGCCCGGTCCTTCTGCGCAACCTCCGTCCCGATCTGCCCTTTAC 988

     95 ............ProTrpValLysProTrpLeuAsnLysProArgLysPr 107
                   |||   :::||||||         ||||||   |||||  
    989 CCCGCCCTTACCCCCTCTATAAAGCCC.........AAACCTCCTAAGCC 1029

    107 o.....................GlyProArgIleLeuAlaLeuGlyGluL 117
        |                     ||||||  |||   |||  |||      
   1030 CCAGGTTCTCCCTGATAGCGGCGGACCTCTCATTGACCTTCTCACAGAGG 1079

    117 ysAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuPro 133
                      :::      ::::::||||||||||||          
   1080 ACCCCCCGCCGTACGGA......GCACAACCTTCCTCCTCT......GCC 1117

    134 ArgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProTh 150
        |||:::            :::   |||:::|||                 
   1118 AGAGAGGACGATGAAAAAGAGGCGGCCGCCACC.................. 1150

    150 rProLeuSerSerThrGlyCysCysGluGlyThrSerAlaProProGlyA 167
                                        :::::::::||||||   :
   1151 ..........................TCCGAGGTTCCCCCCCCTT 1169

    167 laProValValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAla 183
        ::|||:::|||           :::      |||:::||||||        
   1170 CTCCCATGGTG...........TCTCGACTGCGGGGAAGGAGGGACCCT 1207

    184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGl 200
                  :::   :::         :::|||||||||             
   1208 CCCGCAGCGGACTCCACCTCCTCCCAGGCATTCCCACTCCGTATG..... 1252

    200 yProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheS 217
                     ||||||:::   |||   ||||||||||||||||||  
   1253 .............GGGGGAGATGGCCAG...CTTCAGTACTGGCCGTTTT 1286

    217 erSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
        |||||:::|||           |||:::|||:::|||   |||||||||  
   1287 CCTCCTCTGACTT.........AAAAATAATAACCCTTCCTTTTCTGAA 1326

    234 AspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGl 250
        ||||||   :::|||||||:::|||:::||||||:::::::::::|||||  
   1327 GACCCAGGTAAATTGACGGCTTTGATTGAGTCCGTCCTCATCACCCACCA 1376
```

```
 250 nProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrThrG 267
     ||||||||||||||||||||||||||||||   ||||||:::|||   |
1377 GCCCACCTGGGACGACTGTCAGCAGTTGTTGGGGACCTTGCTGACCGGAG 1426

 267 luGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAla 283
     |||||:::::::|||:::|||||||||||||||||||   |||   |||
1427 AAGAAAAGCAGCGGGTGCTCCTAGAGGCTAGAAAGGCAGTCCGGGGCAAT 1476

 284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLe 300
     ||||||||||||||||||||   ||||||:::::   :::||||||||
1477 GATGGACGCCCCACTCAGTTGCCTAATGAAGTCAATGCTGCTTTTCCCCT 1526

 300 uThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuL 317
     |   ||||||   |||||||||:::|||:::||||||||||:::   |||
1527 TGAACGCCCCGATTGGGATTACACCACTACAGAAGGTAGGAACCACCTAG 1576

 317 ysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
       :::|||||||||   |||:::|||||||||:::::::|||:::|||
1577 TCCTCTATCGCCAGTTGCTCTTAGCGGGTCTCCAAAACGCGGGCAGAAGC 1626

 334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluPr 350
     ||||||||||||||||||||:::   :::   ||||||||||||||||
1627 CCCACCAATTTGGCCAAGGTAAAAGGGATAACCCAGGGACCTAATGAGTC 1676

 350 oProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrP 367
      ||||||:::|||||||||||||||   ||||||:::||||||:::||||
1677 TCCCTCAGCCTTTTTAGAGAGACTCAAGGAGGCATATCGCAGGTACACTC 1726

 367 roPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPhe 383
     ||:::||||||   :::   :::|||:::::::::|||:::::::::|||
1727 CTTATGACCCTGAGGACCCAGGGCAAGAAACCAATGTGTCTATGTCATTC 1776

 384 IleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGl 400
     |||   |||||||||   ||||||   :::||||||:::||||||:::
1777 ATCTGGCAGTCTGCCCCGGATATCGGGCGAAAGTTAGAGCGGTTAAAAGA 1826

 400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValT 417
      |||::::::      |||   |||||||||||||||||||||||:::::
1827 TTTAAAGAGCAAGACCTTAGGAGACTTAGTGAGGGAAGCTGAAAAGATCT 1876

 417 yrTyrArgArgGluThrGluGluGluLysGluGlnArgLysGluLysGlu 433
     ::   :::|||||||||   ||||||:::|||:::|||   ::::::|||
1877 TTAATAAGCGAGAAACCCCGGAAGAAAGAGAGGAACGTATCAGGAGAGAA 1926

 434 ArgGluGluArgGluGluArg..............................  440
        ||||||:::|||||||||
1927 ACAGAGGAAAAGGAAGAACGCCGTAGGGCAGAGGATGAGCAGAGAGAGAA 1976

 441 ....ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaV 456
         ||||||||||||:::::::::::::::::::|||:::||||||:::|
1977 AGAAAGGGACCGCAGGAGACATAGAGAGATGAGCAAGCTCTTGGCCACTG 2026

 456 alValGluGlyLysSerSerArgGluArgGluArgAspPheArgLysIle 472
     |||||:::|||:::
2027 TAGTTAGTGGACAG..................................... 2040

 473 ArgSerGlyProArgGlnSerGlyAsnLeuGlyAsnArgThrProLeuAs 489
     |||:::|||   ||||||:::|||:::      :::|||      |||||
2041 AGACAGGGT...AGACAGGGGGGAGAG...CGAAGGAGGCCCCAACTTGA 2084

 489 pLysAspGlnCysAlaTyrCysLysGluLysGlyHisTrpAlaArgAsnC 506
     ||||||||||||||||||||||||||||||||||||||||||||:::::|
2085 TAAGGACCAATGCGCCTACTGCAAAGAAAAAGGACACTGGGCTAAGGACT 2134

 506 ysProLysLys......GlyAsnLysGlyProLys 515
     ||||||||||      |||   :::||||||:::
2135 GCCCAAAGAAGCCACGAGGACCCCGAGGACCGAGG 2169

seq_name: gb_gss:BH063533

seq_documentation_block:
LOCUS       BH063533      801 bp    DNA             GSS       18-JUL-2001

DEFINITION  RPCI-24-370H8.TJ RPCI-24 Mus musculus genomic clone RPCI-24-370H8,
            DNA sequence.
ACCESSION   BH063533
VERSION     BH063533.1  GI:14878283
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-370H8.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 370  row: H  column: 8
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-370H8"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      227 a    193 c    217 g    164 t
ORIGIN


alignment_scores:
            Quality:  991.00            Length:    271
              Ratio:  4.164               Gaps:      1
 Percent Similarity:  87.823   Percent Identity: 71.587

alignment_block:
 US-09-171-553B-4 x BH063533   ..

 Align seg 1/1  to: BH063533  from: 1  to: 801

     215 ProPheSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPh 231
         ||||||||||||:::|||   |||:::|||||||||||||||||||||||
       1 CCTTTTCCTCTTCTGATTA.TATACTTGGAAAACTAACCACCCTCCCTT 49

     231 eSerGluAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheS 248
         |||||||:::|||:::   ||||||||||||:::|||||||||||||||
      50 CTCAGAGAACCCCTCTGGGCTTACTGGGCTCCTTGAGTCACTTATGTTCT 99

     248 erHisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPhe 264
         |||||||||||||||||||||||||||||||||||||||||:::||||||
     100 CCCATCAACCCACTTGGGATGATTGTCAGCAGCTTTTGCAGGTTCTTTTC 149

     265 ThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnValPr 281
         ||||||||||||||||||||||||||:::|||||||||||||||||||
```

```
     150 ACCACAGAAGAAAGAGAAAGAATCCTGATGGAGGCGAGAAAAAATGTTCT 199

     281 oGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyP 298
         |||   ||||||   ||||||   |||   |||   :::|||   :::|
     200 GGGAGAGGACGGCACTCCCACTGCCCTCCCTAACCTCGTGGACGAGGCTT 249

     298 heProLeuThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGlu 314
         ||||||||:::||||||:::||||||||||||||||||||||||||||||
     250 TCCCCTTGAACCGCCCCAACTGGGACTACAACACCGCGGAAGGTAGGGGA 299

     315 SerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSe 331
             |||   :::||||||:::::::||||||||||||||||||||||::
     300 CGCCTCCTTGTCTATCGCCGGACTCTAGTGGCAGGTCTCAGAGGAGCCAC 349

     331 rArgArgProThrAsnLeuAlaLysValArgGluValMetGlnGlyProA 348
         :|||||||||||||||||||||||||||||||||||:::||||||   :
     350 TAGACGGCCCACCAATTTGGCTAAGGTAAGAGAGGTCTTGCAGGGGCAGA 399

     348 snGluProProSerValPheLeuGluArgLeuMetGluAlaPheArgArg 364
         ::|||||||||||||||||||:::|||||||||||||||:::|||||||
     400 CTGAACCACCCTCAGTCTTCCTTGATCGTCTCATGGAGGCATATAGGAGG 449

     365 PheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLe 381
         :::|||||||||||||||:::||||||:::|||||||||:::||||||::
     450 TACACCCCTTTTGACCCCTCGTCAGAGGGGCAGAAAGCCGCTGTAGCCAT 499

     381 uAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgL 398
         :|||||||||||||||||||   ||||||:::||||||||||||||||||
     500 GGCCTTCATTGGTCAGTCCGCTCCCGACATTAAGAAAAAGCTGCAAAGGC 549

     398 euGluGlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGlu 414
         ||||||||||||||:::      |||:::||||||||||:::|||||||||
     550 TGGAGGGGCTCCAAGATCATACGCTCCAAGATTTAGTAAAAGAAGCAGAG 599

     415 LysValTyrTyrArgArgGluThrGluGluGluLysGluGlnArgLysGl 431
         ||||||||:::::::|||||||||:::||||||:::::::::|||:::::
     600 AAAGTCTATCATAAGAGGGAAACCAGAGAAGAGAGGCAGGAGAGAGAGAA 649

     431 uLysGluArgGluGluArgGluGluArgArgAspArgArgGlnGluLysA 448
         :|||||||   ||||||||||||:::||||||||||||||||||||:::|
     650 GAAAGAAATGGAGGAGAGGGAAAATAGACGGGATCGGCGTCAGGAGAGAA 699

     448 snLeuThrLysIleLeuAlaAlaValValGluGlyLysSerSerArgGlu 464
         |||||:::|||||||||||||||||||:::   :::
     700 ATTTGAGTAAAATTTTGGCCGCAGTTGTGAATGATAGA..... 737

     465 ArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAs 481
                                 :::||||||      :::::::|||
     738 .....CAGTCAGGGAAAGGTAAAACAGGACT 763

     481 nLeuGlyAsnArg 485
         ||||||||||||
     764 TCTGGGCAACAGG 776

seq_name: gb_gss:AZ318021

seq_documentation_block:
LOCUS       AZ318021      729 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0036004R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0036004 R, DNA sequence.
ACCESSION   AZ318021
VERSION     AZ318021.1  GI:10367383
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0036  row: O  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 729.
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0036O04"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      196 a    193 c    182 g    158 t
ORIGIN


alignment_scores:
           Quality:  946.00            Length:    239
             Ratio:  4.359               Gaps:      2
 Percent Similarity:  90.795   Percent Identity: 75.314

alignment_block:
 US-09-171-553B-4 x AZ318021   ..

 Align seg 1/1  to: AZ318021  from: 1  to: 729

     194 LeuProLeuArgThrTyr...GlyProProMetProGly...GlyGlnLe 208
         |||   ||||||::::::   |||   |||   ||||||   ::::::||
      11 CTTGCCCTTAGAGCACATGTGGGAGGGCCAGCGCCAGGACCTAATGATCT 60

     208 uGlnProLeuGlnTyrTrpProPheSerSerAlaAspLeuTyrAsnTrpL 225
         |   |||||||||||||||||||||||:::|||||||||||||||||
      61 CATTCCTTTACAATACTGGCCTTTTTCCTCTTCTGATTTATATAATTGGA 110

     225 ysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGlyLeu 241
         ||||||||||||||||||||||||:::|||:::   ||||||||||||
     111 AAACTAACCACCCTCCCTTCTCAGAGAACCCCTCTGGGCTTACTGGGCTC 160

     242 ValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGl 258
         :::||||||||||||||||||||||||||||||||||||||||||||||
     161 CTTGAGTCACTTATGTTCTCCCATCAACCCACTTGGGATGATTGTCAGCA 210
```

```
     258 nLeuLeuGlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuG 275
         ||||||||||:::|||||||||||||||||||||||||||||||||||:::|
     211 GCTTTTGCAGGTTCTTTTCACCACAGAAGAAAGAGAAAGAATCCTGATGG 260

     275 luAlaArgLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGln 291
         ||||||||||||||||   |||   |||      ||||||   |||
     261 AGGCGAGAAAAAATGTTCTAGGAGAGGACGACACACCCACTGCCCTCCCT 310

     292 AsnGluIleAspMetGlyPheProLeuThrArgProGlyTrpAspTyrAs 308
         |||   :::|||   :::|||||||||:::||||||:::||||||||||
     311 AACCTCGTGGACGAGGCTTTCCCCTTGAACCGCCCCAACTGGGACTACAA 360

     308 nThrAlaGluGlyArgGluSerLeuLysIleTyrArgGlnAlaLeuValA 325
         ||||||||||||||||      |||   :::||||||::::::|||||||
     361 CACCGCGGAAGGTAGGGGACGCCTCCTTGTCTATCGCCGGACTCTAGTGG 410

     325 laGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArg 341
         ||||||||||||||||:::||||||||||||||||||||||||||||||
     411 CAGGTCTCAGAGGAGCCGCTAGACGGCCCACCAATTTGGCTAAGGTAAGA 460

     342 GluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLe 358
         ||||||:::||||||   :::||||||||||||||||||||||||||||||
     461 GAGGTCTTGCAGGGGCAGACTGAACCACCCTCAGTCTTCCTTGAGCGTCT 510

     358 uMetGluAlaPheArgArgPheThrProPheAspProThrSerGluAlaG 375
         |||||||||:::::|||:::|||||||||||||||:::||||||:::|
     511 CATGGAGGCATATAAGAGGTACACCCCTTTTGACCCCTCGTCAGAGGGGC 560

     375 lnLysAlaSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIle 391
         ||||||||:::|||||:::|||||||||||||||||||   ||||||
     561 AAAAAGCCGCTGTAGCCATGGCCTTCATTGGTCAGTCCGCTCCCGACATT 610

     392 ArgLysLysLeuGlnArgLeuGluGlyLeuGlnGluAlaGluLeuArgAs 408
         :::||||||||||||||||||||||||||:::      |||:::||
     611 AAGAAAAAGCTGCAAAGGCTGGAGGGGCTCCAAGATCATACGCTCCAAGA 660

     408 pLeuValArgGluAlaGluLysValTyrTyrArgArgGluThrGluGluG 425
         |||||||:::|||||||||||||||||:::::::|||||||||||||||
     661 TTTAGTAAAAGAAGCAGAGAAAGTCTATCATAAGAGGGAAACAGAAGAAG 710

     425 luLysGluGlnArgLys 430
         ||:::::::::|||:::
     711 AGAGGCAGGAGAGAGAG 727

seq_name: gb_gss:BH121259

seq_documentation_block:
LOCUS       BH121259      679 bp    DNA             GSS       19-JUL-2001
DEFINITION  RPCI-24-374G11.TV RPCI-24 Mus musculus genomic clone RPCI-24-374G11
            , DNA sequence.
ACCESSION   BH121259
VERSION     BH121259.1  GI:14964771
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 679)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-374G11.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 374  row: G  column: 11
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .679
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="RPCI-24-374G11"
                      /clone_lib="RPCI-24"
                      /sex="Male"
                      /cell_type="Spleen/Brain"
                      /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                      RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                      library was  cloned in the pTARBAC1 cloning vector at the
                      BamH1 sites using MboI  partially digested male C57BL/6J
                      DNA."
BASE COUNT      194 a    167 c    174 g    144 t
ORIGIN


alignment_scores:
            Quality:  916.00              Length:    226
              Ratio:  4.425                 Gaps:      1
 Percent Similarity:  91.593   Percent Identity: 76.991

alignment_block:
 US-09-171-553B-4 x BH121259   ..

 Align seg 1/1  to: BH121259  from: 1  to: 679

     214 TrpProPheSerSerAlaAspLeuTyrAsnTrp.LysThrAsnHisProP 230
         ||||||||||||||:::|||||||||||||| ||||||||||||||||
       1 TGGCCTTTTTCCTCTTCTGATTTATACAATTGGGAAAACTAACCACCCTC 50

     230 roPheSerGluAspProGlnArgLeuThrGlyLeuValGluSerLeuMet 246
         ||||||||||:::|||:::   ||||||||||||:::|||||||||||||
      51 CCTTCTCAGAGAACCCCTCTGGACTTACTGGGCTCCTTGAGTCACTTATG 100

     247 PheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLe 263
         ||||||||||||||||||||||||||||||||||||||||||||||:::||
     101 TTCTCCCATCAACCCACTTGGGATGATTGTCAGCAGCTTTTGCAGGTTCT 150

     263 uPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnV 280
         ||||||||||||||||||||||||||||||:::|||||||||||||||||
     151 TTTTACCACAGAAGAAAGAGAAAGAATCCTGATGGAGGCGAGAAAAAATG 200

     280 alProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMet 296
         ||   |||   ||||||   ||||||   |||   |||   :::|||
     201 TTCTGGGAGAGGACGGCACACCCACTGCCCTCCCTAACCTCGTGGACGAG 250

     297 GlyPheProLeuThrArgProGlyTrpAspTyrAsnThrAlaGluGlyAr 313
         :::||||||||||:::::::|||:::||||||||||||||||||||||||
     251 GCTTTTCCCTTGAACCACCCCAACTGGGACTACAACACCGCGGAAGGTAG 300

     313 gGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyA 330
         |      |||   :::||||||::::::|||||||||||||||||||||||
     301 GGGACGCCTCCTTGTCTATCGCCGGACTCTAGTGGCAGGTCTCAGAGGAG 350

     330 laSerArgArgProThrAsnLeuAlaLysValArgGluValMetGlnGly 346
         ||:::|||||||||||||||||||||||||||||||||||:::||||||
     351 CCGCTAGACGGCCCACCAATTTGGCTAAGGTAAGAGAGGTCTTGCAAGGG 400

     347 ProAsnGluProProSerValPheLeuGluArgLeuMetGluAlaPheAr 363
         :::|||||||||||||||||||||||||||||||||||||||||||:::||
     401 CAGACTGAACCACCATCAGTCTTCCTTGAGCGTCTAATGGAGGCATATAG 450

     363 gArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValA 380
         ||||:::|||||||||||||||   ||||||:::|||:::|||:::||||
     451 GAGATACACCCCTTTTGACCCCTTGTCAGAGGGGCAGAGAGCCGCTGTAG 500
```

```
     380 laLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGln 396
         ||:::|||||||||||||||||||||   ||||||:::|||||||||||||
     501 CCATGGCCTTCATTGGTCAGTCCGCTCCCGACATTAAGAAAAAGCTACAA 550

     397 ArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAl 413
         ||||||||||||||||||:::      |||:::|||||||||:::||||||
     551 AGGCTGGAGGGGCTCCAAGATCATACGCTCCAAGATTTAGTAAAAGAAGC 600

     413 aGluLysValTyrTyrArgArgGluThrGluGluGluLysGluGlnArgL 430
         ||||||||||||:::::|||||||||||||||||:::   :::|||:
     601 AGAGAAAGTCTATCATAAGAGGGAAACAGAAGAAGAGAGGCTGGAGAGAG 650

     430 ysGluLysGluArgGluGluArgGlu 438
         :::::||||||   ||||||||||||
     651 AGAAGAAAGAAATAGAGGAGAGGGAA 676

seq_name: gb_gss:AZ967275

seq_documentation_block:
LOCUS       AZ967275      646 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0238G10F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0238G10 F, DNA sequence.
ACCESSION   AZ967275
VERSION     AZ967275.1  GI:13838502
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0238  row: G  column: 10
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 646.
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0238G10"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      190 a    150 c    182 g    124 t
ORIGIN


alignment_scores:
           Quality:  861.00            Length:    214
             Ratio:  4.348               Gaps:      0
 Percent Similarity:  92.523   Percent Identity: 76.636

alignment_block:
 US-09-171-553B-4 x AZ967275   ..

 Align seg 1/1  to: AZ967275  from: 1  to: 646

     247 PheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLe 263
         ||||||||||||||||||||||||||||||||||||||||||||||||:::||
       1 TTCTCCCATCAACCCACTTGGGATGATTGTCAGCAGCTTTTGCAGGTTCT 50

     263 uPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnV 280
         ||||||||||||||||||||||||||||||||:::|||||||||||||||
      51 TTTCACCACAGAAGAAAGAGAAAGAATCCTGATGGAGGCGAGAAAAAATG 100

     280 alProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMet 296
         ||   |||   ||||||   ||||||   |||   |||   :::|||
     101 TTCTAGGAGAGGACGGCACACCCACTGCCCTCCCTAATCTCGTGGACGAG 150

     297 GlyPheProLeuThrArgProGlyTrpAspTyrAsnThrAlaGluGlyAr 313
         :::|||||||||:::||||||:::||||||||||||||||||||||||||
     151 GCTTTCCCCTTGAACCGCCCCAACTGGGACTACAACACCGCGGAAGGTAG 200

     313 gGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyA 330
         |       |||   :::||||||:::::|||||||||||||||||||||||
     201 GGGACGCCTCCTTGTCTATCGCCGGACTCTAGTGGCAGGTCTCAGAGGGG 250

     330 laSerArgArgProThrAsnLeuAlaLysValArgGluValMetGlnGly 346
         ||:::|||||||||||||||||||||||||||||||||||||:::||||||
     251 CCGCTAGACGGCCCACCAATTTGGCTAAGGTAAGAGAGGTCTTGCAGGGG 300

     347 ProAsnGluProProSerValPheLeuGluArgLeuMetGluAlaPheAr 363
         :::|||||||||||||||||||||||||||||||||||||||||||:::||
     301 CAGACTGAACCACCCTCAGTCTTCCTTGAGCGTCTCATGGAGGCATATAG 350

     363 gArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValA 380
         ||||:::|||||||||||||||||:::||||||:::||||||||:::||||
     351 GAGGTACACCCCTTTTGACCCCTCGTCAGAGGGGCAGAAAGCCGCTGTAG 400

     380 laLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGln 396
         ||:::|||||||||||||||||||   ||||||:::|||||||||||||
     401 CCATGGCCTTCATTGGTCAGTCCGCTCCCGACATTAAGAAAAAGCTGCAA 450

     397 ArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAl 413
         ||||||||||||||||||:::      |||:::|||||||||:::||||||
     451 AGGCTGGAGGGGCTCCAAGATCATACGCTCCAAGATTTAGTAAAAGAAGC 500

     413 aGluLysValTyrTyrArgArgGluThrGluGluGluLysGluGlnArgL 430
         ||||||||||||:::::|||||||||||||||||||||||||||||||:
     501 AGAGAAAGTCTATCATAAGAGGGAAACAGAAGAAGAGAGGCAGGAGAGAG 550

     430 ysGluLysGluArgGluGluArgGluGluArgArgAspArgArgGlnGlu 446
         :::::||||||   ||||||||||||:::||||||||||||||||||||
     551 AGAAGAAAGAAATGGAGGAGAGGGAAAATAGACGGGATCGCCGTCAGGAG 600

     447 LysAsnLeuThrLysIleLeuAlaAlaValValGluGlyLys 460
         :::||||||:::|||||||||||||||||||||:::   :::
     601 AGAAATTTGAGTAAAATTTTGGCCGCAGTTGTAAATGATAGA 642

seq_name: gb_gss:BH124013
```

```
seq_documentation_block:
LOCUS       BH124013      644 bp    DNA             GSS       19-JUL-2001
DEFINITION  RPCI-24-230P16.TV RPCI-24 Mus musculus genomic clone RPCI-24-230P16
            , DNA sequence.
ACCESSION   BH124013
VERSION     BH124013.1  GI:14967525
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 644)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-230P16.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 230  row: P  column: 16
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .644
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-230P16"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      176 a    162 c    169 g    137 t
ORIGIN


alignment_scores:
            Quality:  846.00            Length:    213
              Ratio:  4.383               Gaps:      0
 Percent Similarity:  90.610   Percent Identity: 75.117

alignment_block:
 US-09-171-553B-4 x BH124013   ..

 Align seg 1/1  to: BH124013  from: 1  to: 644

     223 AsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuTh 239
         ||||||||||||||||||||||||||||||||||:::|||:::   |||||
       3 AATTGGAAAACTAACCACCCTCCCTTCTCAGAGAACCCCTCTGGACTTAC 52

     239 rGlyLeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspC 256
         |||||||:::|||||||||||||||||||||||||||||||||||||||||
      53 TGGGCTCCTTGAGTCACTTATGTTCTCCCATCAACCCACTTGGGATGATT 102

     256 ysGlnGlnLeuLeuGlnThrLeuPheThrThrGluGluArgGluArgIle 272
         ||||||||||||||||:::|||||||||||||||||||||||||||||||
     103 GTCAGCAGCTTTTGCAGGTTCTTTTTACCACAGAAGAAAGAGAAAGAATC 152

     273 LeuLeuGluAlaArgLysAsnValProGlyAlaAspGlyArgProThrGl 289
         |||:::||||||          |||   |||   ||||||   ||||||
     153 CTGATGGAGGCGATTTTTTTTGTTCTGGGAGAGGACGGCACACCCACTGC 202

     289 nLeuGlnAsnGluIleAspMetGlyPheProLeuThrArgProGlyTrpA 306
          |||   |||   :::|||   :::||||||||||:::||||||:::||||
     203 CCTCCCTAACCTCGTTGACGAGGCTTTTCCCTTGAACCGCCCCAACTGGG 252

     306 spTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArgGlnAla 322
         ||||||||||||||||||||||||||      |||   :::||||||:::::
     253 ACTACAACACCGCGGAAGGTAGGGGACGCCTCCTTGTCTATCGCCGGACT 302

     323 LeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLy 339
         ||||||||||||||||||||||||:::||||||||||||||||||||||||
     303 CTAGTGGCAGGTCTCAGAGGAGCCGCTAGACGGCCCACCAATTTGGCTAA 352

     339 sValArgGluValMetGlnGlyProAsnGluProProSerValPheLeuG 356
         ||||||||||||:::||||||   :::|||||||||||||||||||||||
     353 GGTAAGAGAGGTCTTGCAGGGGCAGACTGAACCACCCTCAGTCTTCCTTG 402

     356 luArgLeuMetGluAlaPheArgArgPheThrProPheAspProThrSer 372
         ||||||||||||||||:::||||||:::|||||||||||||||   |||
     403 AGCGTCTAATGGAGGCATATAGGAGATACACCCCTTTTGACCCCTTGTCA 452

     373 GluAlaGlnLysAlaSerValAlaLeuAlaPheIleGlyGlnSerAlaLe 389
         |||:::|||:::|||:::||||||::||||||||||||||||||||||||
     453 GAGGGGCAGAGAGCCGCTGTAGCCATGGCCTTCATTGGTCAATCCGCTCC 502

     389 uAspIleArgLysLysLeuGlnArgLeuGluGlyLeuGlnGluAlaGluL 406
          ||||||:::|||||||||||||||||||||||||||||||:::     |
     503 CGACATTAAGAAAAAGCTGCAGAGGCTGGAGGGGCTCCAAGATCATACGC 552

     406 euArgAspLeuValArgGluAlaGluLysValTyrTyrArgArgGluThr 422
         ||:::||||||||||:::|||||||||||||||||||::::::|||||||||
     553 TCCAAGATTTAGTAAAAGAAGCAGAGAAAGTCTATCATAAAAGGGAAACA 602

     423 GluGluGluLysGluGlnArgLysGluLysGluArgGlu 435
         :::||||||:::::::::|||::::::||||||   |||
     603 CGAGAAGAGAGGCAGGAGAGAGAGAAGAAAGAAATGGAG 641

seq_name: gb_gss:AZ420279

seq_documentation_block:
LOCUS       AZ420279      617 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0198K05F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0198K05 F, DNA sequence.
ACCESSION   AZ420279
VERSION     AZ420279.1  GI:10544292
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0198  row: K  column: 05
            Seq primer: CGTTGTAAAACGACGGCCAGT
```

```
           Class: plasmid ends
           High quality sequence stop: 617.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0198K05"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      138 a    158 c    155 g    166 t
ORIGIN


alignment_scores:
            Quality:  845.50              Length:    212
              Ratio:  4.450                 Gaps:      1
 Percent Similarity:  89.623   Percent Identity: 76.887

alignment_block:
 US-09-171-553B-4 x AZ420279/rev   ..

 Align seg 1/1  to reverse of: AZ420279  from: 1  to: 617

     219 AlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspPr 235
         :::|||||||||||||||||||||||||||||   ||||||||||||:::||
     615 TCGGATCTTTACAATTGGAAAACTAATCACCACCCTTTCTCTGAAAACCC 566

     235 oGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGlnProT 252
         |:::   ||||||||||||:::|||||||||||||||||||||||||||||
     565 CTCAGGGTTGACTGGACTCCTAGAGTCCCTGATGTTTTCCCACCAGCCCA 516

     252 hrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrThrGluGlu 268
         |||||||||||||||||||||||||||||:::||||||||||||||||||
     515 CTTGGGATGACTGTCAACAGCTTTTACAGGTCCTGTTCACAACGGAAGAA 466

     269 ArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAlaAspGl 285
         |||||||||||||||||||||||||||||||||||   |||   |||||
     465 CGGGAAAGAATTCTTCTGGAGGCTAGAAAAAATGTCTTGGGGAACGATGG 416

     285 yArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeuThrA 302
         |   ||||||:::|||   |||   ||||||   :::|||||||||||||
     415 GACTCCCACCAACCTTCCTAACCTTATTGATGAAGCTTTTCCTTTGACCC 366

     302 rgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIle 318
         |||||:::|||||||||||||||||||||||||||||||   |||   :::
     365 GCCCAAACTGGGACTACAATACAGCAGAAGGTAGGGAGCGTCTCCTGGTC 316

     319 TyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProTh 335
         ||||||::::::|||||||||||||||||:::||||||:::||||||||||
     315 TATCGCCGGACTCTAGTGGCGGGTCTCAAAGGGGCTGCACGACGGCCCAC 266

     335 rAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluProProS 352
         |||||||||||||||||||||||||:::|||||||||   ||||||||||
     265 CAATCTGGCAAAGGTAAGAGAGGTCTTGCAGGGACCAGCCGAGCCACCCT 216

     352 erValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrProPhe 368
         |||||||||||||||||||||||||||||||:::|||||||:::|||||||||
     215 CGGTCTTTCTAGAACGGCTAATGGAGGCCTATAGGAGGTATACCCCGTTC 166

     369 AspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPheIleGl 385
         ||||||:::||||||:::|||:::|||:::||||||:::||||||||||||
     165 GATCCTTCATCGGAGGGGCAGCAAGCAGCTGTTGCTATGGCTTTCATTGG 116

     385 yGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGlyLeuG 402
         ||||||||||   |||||||:::||||||||||||||||||||||||||||
     115 CCAGTCAGCTCCCGATATAAAAAAGAAGCTGCAAAGACTCGAGGGCCTCC 66

     402 lnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyr 418
         ||:::      |||:::||||||||||:::||||||||||||||||:::
      65 AAGATTATACTTTGCAGGATCTAGTGAAGGAAGCAGAGAAAGTGTTTCTA 16

     419 ArgArgGluThrGluGluGluLysGluGlnArgLys 430
         |||                    |||:::||||||
      15 AGA.....................GAGAGACGGAAG 1

seq_name: gb_gss:AZ986580

seq_documentation_block:
LOCUS       AZ986580      677 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0268P18R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0268P18 R, DNA sequence.
ACCESSION   AZ986580
VERSION     AZ986580.1  GI:13857807
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0268  row: P  column: 18
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 677.
FEATURES             Location/Qualifiers
     source          1. .677
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0268P18"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
```

polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 164 a 191 c 172 g 150 t
ORIGIN

alignment_scores:
Quality: 828.50 Length: 224
Ratio: 4.249 Gaps: 3
Percent Similarity: 87.054 Percent Identity: 71.875

alignment_block:
US-09-171-553B-4 x AZ986580 ..

Align seg 1/1 to: AZ986580 from: 1 to: 677

```
    179 SerArgArgGlyAlaThrProGluArg.........ThrAspGluIleAl 192
        ||||||||||||   :::|||   :::            |||:::   ::
      6 TCGCGCCGAGGCCAAAGCCCCGGGGAGGAAGGGGGGCCAGATTCTACAGT 55

    192 aIleLeuProLeuArgThrTyr...GlyProProMetProGly...GlyG 207
        :   ||||||||||||:::::   |||   |||   ||||||   ::::
     56 TGCCTTGCCACTTAGAGCACATGTGGGAGGGCCAGCGCCAGGACCTAATG 105

    207 lnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAspLeuTyrAsn 223
        ::|||   ||||||||||||||||||||||||||:::|||||||||||||
    106 ATCTCATTCCTTTACAATACTGGCCTTTTTCCTCTTCTGATTTATATAAT 155

    224 TrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGl 240
        ||||||||||||||||||||||||||||||:::|||:::   ||||||||
    156 TGGAAAACTAACCACCCTCCCTTCTCAGAGAACCCCTCTGGGCTTACTGG 205

    240 yLeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysG 257
        ||||:::||||||||||||||||||||||||||||||||||   ||||
    206 GCTCCTTGAGTCACTTATGTTCTCCCATCAACCCACTTGGGATGGTTGTC 255

    257 lnGlnLeuLeuGlnThrLeuPheThrThrGluGluArgGluArgIleLeu 273
        |||||   ||||||:::|||||||||||||||||||||||||||||||||
    256 AGCAGCCTTTGCAGGTTCTTTTCACCACAGAAGAAAGAGAAAGAATCCTG 305

    274 LeuGluAlaArgLysAsnValProGlyAlaAspGlyArgProThrGlnLe 290
        :::||||||||||||||||   |||   ||||||   ||||||   ||
    306 ATGGAGGCGAGAAAAAATGTTCTGGGAGAGGACGGCACACCCACTGCCCT 355

    290 uGlnAsnGluIleAspMetGlyPheProLeuThrArgProGlyTrpAspT 307
        |   |||   :::|||   :::|||||||||:::||||||:::|||||||
    356 CCCTAACCTCGTTGACGAGGCTTTCCCCTTGAACCGCCCCAACTGGGACT 405

    307 yrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArgGlnAlaLeu 323
        ||||||||:::|||||||||       |||   :::||||||:::::|||
    406 ACAACACCACGGAAGGTAGGGGACGCCTCCTTGTCTATCGCCGGACTCTA 455

    324 ValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVa 340
        ||||||||||||||||||||:::|||:::|||||||||||||||||||||
    456 GTGGCAGGTCTCAGAGGAGCCGCTAGACAGCCCACCAATTTGGCTAAGGT 505

    340 lArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluA 357
        ||||||||||:::||||||   :::|||||||||||||||||||||||||
    506 AAGAGAGGTCTTGCAGGGGCAGACTGAACCACCCTCAGTCTTCCTTGAGC 555

    357 rgLeuMetGluAlaPheArgArgPheThrProPheAspProThrSerGlu 373
        ||||||||||||||:::||||||:::|||||||||||||||   ||||||
    556 GTCTCATGGAGGCATATAGGAGGTACACCCCTTTTGACCCCTTGTCAGAG 605

    374 AlaGlnLysAlaSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAs 390
        :::|||:::|||:::||||||:::|||||||||||||||||||||   ::
    606 GGGCAGAGAGCCGCTGTAGCCATGGCCTTCATTGGTCAGTCCGCTCCCAA 655

    390 pIleArgLysLysLeuGlnArg 397
        :|||:::|||||||||||||||
    656 CATTAAGAAAAAGCTGCAAAGG 677
```

seq_name: gb_gss:AZ862103

seq_documentation_block:

LOCUS AZ862103 677 bp DNA GSS 21-FEB-2001
DEFINITION 2M0169D03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0169D03 R, DNA sequence.
ACCESSION AZ862103
VERSION AZ862103.1 GI:13059087
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 677)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0169 row: D column: 03
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 677.

FEATURES Location/Qualifiers
source 1. .677
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0169D03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 159 a 173 c 181 g 164 t
ORIGIN

```
alignment_scores:
            Quality:  827.00              Length:    223
              Ratio:  4.241                 Gaps:      3
 Percent Similarity:  87.444   Percent Identity: 73.991

alignment_block:
 US-09-171-553B-4 x AZ862103/rev   ..

 Align seg 1/1  to reverse of: AZ862103  from: 1  to: 677

     189 AspGluIleAlaIleLeuProLeuArgThrTyr...GlyProProMetPr 204
         |||:::   :::   ||||||||||||:::::   |||   |||   ||
     660 GATTCAACAGTTGCCTTGCCACTTAGAGCACATGTGGGAGGGCCAGCGCC 611

     204 oGly...GlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaA 220
         ||||   :::::::|||   |||||||||||||||||||||||||||:::|
     610 AGGACCTAATGATCTCATTCCTTTACAATACTGGCCTTTTTCCTCTTCTG 561

     220 spLeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGln 236
         |||||||||||||||||||||||||||||||||||||||||||:::|||:::
     560 ATTTATATAATTGGAAAACTAACCACCCTCCCTTCTCAGAGAACCCCTCT 511

     237 ArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGlnProThrTr 253
         |||||||||||||:::|||||||||||||||||||||||||||||||||
     510 GGGCTTACTGGGCTCCTTGAGTCACTTATGTTCTCCCATCAACCCACTTG 461

     253 pAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrThrGluGluArgG 270
         |||||||||||||||||||||||:::|||||||||||||||||||||||
     460 GGATGATTGTCAGCAGCTTTTGCAGGTTCTTTTTACCACAGAAGAAAGAG 411

     270 luArgIleLeuLeuGluAlaArgLysAsnValProGlyAlaAspGlyArg 286
         ||||||||||:::|||||||||||||||||||   |||   ||||||
     410 AAAGAATCCTGATGGAGGCTAGAAAAAATGTTCTAGGAGAGGACGGCACA 361

     287 ProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeuThrArgPr 303
         ||||||   |||   |||   :::|||   :::||||||||||:::||||||
     360 CCCACTGCTCTCCCTAACCTCGTGGATGAGGCTTTCCCCTTGAACCGCCC 311

     303 oGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrA 320
         |:::|||||||||||||||||||||||||||||      |||   :::||||
     310 CAACTGGGACTACAACACCGCGGAAGGTAGGGGACACCTCCTTGTCTATC 261

     320 rgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsn 336
         ||:::::::|||||||||||||||||||||
     260 GCCGGACTCTAGTGGCAGGTCTCAGAGGGGC..................... 230

     337 LeuAlaLysValArgGluValMetGlnGlyProAsnGluProProSerVa 353
         |||||||||||||||||||||||:::|||||||   :::|||||||||||||
     229 CTGGCTAAGGTAAGAGAGGTCTTGCAGGGGCAGACTGAACCACCCTCAGT 180

     353 lPheLeuGluArgLeuMetGluAlaPheArgArgPheThrProPheAspP 370
         |||||||||||||||||||||||||:::|||||||:::|||||||||||||
     179 CTTCCTTGAGCGTCTCATGGAGGCATATAGGAGGTACACCCCTTTTGACC 130

     370 roThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPheIleGlyGln 386
         ||:::|||||:::|||||||||:::|||||||:::|||||||||||||||
     129 CCTCGTCAGAGGGGCAGAAAGCTGCTGTAGCCATGGCCTTCATTGGTCAG 80

     387 SerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGlyLeuGlnGl 403
         ||||||   ||||||:::|||||||||||||||||||||||||||||||::
      79 TCCGCTCCCGACATTAAGAAAAAGCTGCAAAGGCTGGAGGGGCTCCAAGA 30

     403 uAlaGluLeuArgAspLeu 409
         :      |||:::||||||
      29 TCATACGCTCCAAGATTTA 11

seq_name: gb_est2:BG089942

seq_documentation_block:
LOCUS       BG089942      645 bp    mRNA            EST       26-JAN-2001
DEFINITION  mab88d06.x1 NCI_CGAP_SP2 Mus musculus cDNA clone IMAGE:3977506 3'
            similar to TR:Q9WLL1 Q9WLL1 GAG PROTEIN. [1] ;, mRNA sequence.
ACCESSION   BG089942
VERSION     BG089942.1  GI:12572505
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: mab88d06.y1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David Segal Ph.D., Herbert Morse M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov

            MGI:1477538
            Seq primer: -40UP from Gibco
            High quality sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1. .645
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3977506"
                     /clone_lib="NCI_CGAP_SP2"
                     /tissue_type="NK cells (flow-sorted)"
                     /lab_host="DH10B (T1-resistant)"
                     /note="Organ: spleen; Vector: pCMV-SPORT6 (Life
                     Technologies); mRNA made from flow-sorted NK cells, cDNA
                     made by oligo-dT priming.  Directionally cloned. Average
                     insert size 1.5 kb. Primary library, non-amplified.  cDNA
                     Library Preparation: David B. Krizman, Ph.D."
BASE COUNT      141 a    177 c    150 g    176 t      1 others
ORIGIN


alignment_scores:
            Quality:  810.50              Length:    215
              Ratio:  4.311                 Gaps:      1
 Percent Similarity:  87.442   Percent Identity: 75.349

alignment_block:
 US-09-171-553B-4 x BG089942/rev   ..

 Align seg 1/1  to reverse of: BG089942  from: 1  to: 645

     216 PheSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSe 232
         |||      :::||||||||||      |||:::      |||||||||
     642 TTTTTCTTCTGTGACCTTTACAANTGAAAGGTC...ATCCCCCCCTTTTT 596

     232 rGluAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerH 249
          :::      |||   |||||||||||||||||||||||||||:::||||
     595 CAGAAAACTGCAAGGACTCACTGGGTTGGTTGAATCATTAATGTATTCAC 546

     249 isGlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThr 265
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     545 ACCAGCCGACTTGGGATGACTGCCAGCAGCTTCTGCAGACTCTATTCACA 496

     266 ThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGl 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     495 ACCGAGGAGAGAGAGAGGATTCTCCTCGAGGCTCGGAAAAATGTCCGAGA 446

     282 yAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheP 299
              |||||||||:::|||         |||||||||   |||||||
     445 CGAGGCTGGGCGCCCTGTCCAAACTCCAGCTGAGATAGATGAAGGATTTC 396
```

```
     299 roLeuThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSer 315
         |||||||||||||||   ||||||||||||||||||:::|||||||||
     395 CGCTAACCCGGCCCCGATGGGATTATAATACGGCATCAGGTAGGGAACGA 346

     316 LeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerAr 332
         |||:::   ||||||:::::|||||||||||||||||||||||:::||
     345 CTGTCCAATTATCGCCGGGTCCTAGTGGCGGGTCTCAGAGGTGCTGCCCG 296

     332 gArgProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnG 349
         |:::|||||||||||||||||||||||||||||||||||||||   :::|
     295 GCAGCCCACGAATCTGGCCAAGGTAAGAGAGGTTATGCAGGGAGCGACTG 246

     349 luProProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPhe 365
         |||||||||||||||||||||||||||||||||||||:::|||||||:::
     245 AGCCCCCCTCAGTCTTCCTTGAAAGGCTCATGGAGGCTTATAGGAGATAT 196

     366 ThrProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAl 382
         |||||||||||||||||||||||||:::|||:::|||||||||   :::||
     195 ACCCCATTCGACCCCACGTCTGAGGGTCAAAGGGCCTCAGTAATTATGGC 146

     382 aPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuG 399
         ||||||||||||||||||   |||||||||||||||||||||||:::|
     145 CTTCATTGGCCAGTCAGCTCCTGACATTAGGAAGAAGTTACAGCGAATTG 96

     399 luGlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLys 415
         ||||||||||:::      :::|||||||:::|||||||||||||||||
      95 AGGGCTTGCAGGATTACACCATAAGGGATGTAGTTAGAGAGGCAGAGAAA 46

     416 ValTyrTyrArgArgGluThrGluGluGluLysGluGlnArgLys 430
         ||||||:::|||||||||||||||||||:::::|||   :::|||:::
      45 GTGTATCATAGGAGAGAAACAGAAGATAAAAAGTTAGAGAGAGAG 1

seq_name: gb_est1:BE283560

seq_documentation_block:
LOCUS       BE283560      598 bp    mRNA            EST       26-OCT-2000
DEFINITION  601102760F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3495206 5',
            mRNA sequence.
ACCESSION   BE283560
VERSION     BE283560.1  GI:9159895
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8545  row: g  column: 15
            High quality sequence stop: 598.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Mus musculus"
                     /strain="CZECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3495206"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      144 a    164 c    155 g    135 t
ORIGIN


alignment_scores:
            Quality:  809.00             Length:    193
              Ratio:  4.649                Gaps:      0
 Percent Similarity:  90.155   Percent Identity: 79.275

alignment_block:
 US-09-171-553B-4 x BE283560   ..

 Align seg 1/1  to: BE283560  from: 1  to: 598

     201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSe 217
         ||||||         ::::::||||||||||||||||||||||||||||||
      18 CCTCCTGCCGATCCAAATAGTCTACAGCCCCTACAGTATTGGCCTTTTTC 67

     217 rSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGluA 234
         ||||:::|||||||||||||||||:::::|||||||||||||||||||:
      68 CTCTTCTGACCTTTACAACTGGAAAGCTCATCACCCCCCTTTTTCAGAAA 117

     234 spProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGln 250
         ::|||      ||||||||||||||||||||||||||:::|||||||||
     118 ACCCTGCAGGACTCACTGGGTTGGTTGAATCATTAATGTATTCACACCAG 167

     251 ProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrThrGl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     168 CCGACCTGGGATGACTGCCAGCAGCTTCTGCAGACTCTATTCACAACCGA 217

     267 uGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAlaA 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     218 GGAGAGAGAGAGGATTCTCCTCGAGGCTCGGAAAAACGTCCGAGATGAGG 267

     284 spGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300
           |||||||||:::|||         |||||||||   ||||||||||||
     268 CTGGGCGCCCTGTCCAAACTCCAGCTGAGATAGATGAAGGATTTCCGCTA 317

     301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLy 317
         |||||||||   ||||||||||||||||||:::|||||||||   |||::
     318 ACCCGGCCCCGATGGGATTATAATACGGCATCAGGTAGGGAACGACTGTC 367

     317 sIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgP 334
         :   |||||:::::|||||||||||||||||||||||||||:::|||:::|
     368 CAATTATCGCCGGGTCCTAGTGGCGGGTCTCAGAGGTGCTGCCCGGCAGC 417

     334 roThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluPro 350
         |||||||||||||||||||||||||||||||||||||   :::||||||
     418 CCACGAATCTGGCCAAGGTAAGAGAGGTTATGCAGGGAGCGACTGAGCCC 467

     351 ProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrPr 367
         ||||||||||||||||||||||||||||||||:::::|||:::|||||
     468 CCCTCAGTCTTCCTTGAAAGGCTCATGGAGGCTTATAAGAGATATACCCC 517

     367 oPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPheI 384
         ||||||||||||||||||:::|||:::||||||||||   :::|||||||
     518 ATTCGACCCCACGTCTGAGGGTCAAAGGGCCTCAGTAATTATGGCCTTCA 567

     384 leGlyGlnSerAlaLeuAspIleArgLys 393
         |||||||||||||   ||||||||||||
     568 TTGGCCAGTCAGCTCCTGACATTAGGAAG 596

seq_name: gb_gss:AZ650638

seq_documentation_block:
LOCUS       AZ650638      601 bp    DNA             GSS       14-DEC-2000
DEFINITION  1M0520L20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0520L20 R, DNA sequence.
ACCESSION   AZ650638
VERSION     AZ650638.1  GI:11785326
KEYWORDS    GSS.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 601)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0520  row: L  column: 20
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 601.
FEATURES             Location/Qualifiers
     source          1. .601
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0520L20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      175 a    147 c    169 g    110 t
ORIGIN


alignment_scores:
            Quality:  800.00              Length:    199
              Ratio:  4.372                 Gaps:      0
 Percent Similarity:  91.960   Percent Identity: 76.382

alignment_block:
 US-09-171-553B-4 x AZ650638   ..

 Align seg 1/1  to: AZ650638  from: 1  to: 601

     248 SerHisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPh 264
         ||||||||||||||||||||||||||||||||||||||||||||:::|||||
       5 TCCCATCAACCCACTTGGGATGATTGTCAGCAGCTTTTGCAGGTTCTTTT 54

     264 eThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnValP 281
         ||||||||||||||||||||||||||||:::||||||||||||||||||
```

```
      55 CACCACAGAAGAAAGAGAAAGAATCCTGATGGAGGCGAGAAAAAATGTTC 104

     281 roGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGly 297
         |||   ||||||   ||||||   |||   |||   :::|||   :::
     105 TAGGAGAGGACGGCACACCCACTGCCCTCCCTAATCTCGTGGACGAGGCT 154

     298 PheProLeuThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGl 314
         |||||||||:::||||||:::||||||||||||||||||||||||||||
     155 TTCCCCTTGAACCGCCCCAACTGGGACTACAACACCGCGGAAGGTAGGGG 204

     314 uSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaS 331
         |||   :::||||||::::::|||||||||||||||   ||||||:
     205 ACGCCTCCTTGTCTATCGCCGGACTCTAGTGGCAGGTCTCATAGGGGCCG 254

     331 erArgArgProThrAsnLeuAlaLysValArgGluValMetGlnGlyPro 347
         ::|||||||||||||||||||||||||||||||||||||:::||||||
     255 CTAGACGGCCCACCAATTTGGCTAAAGTAAGAGAGGTCTTGCAGGGGCAG 304

     348 AsnGluProProSerValPheLeuGluArgLeuMetGluAlaPheArgAr 364
         :::|||||||||||||||||||||||||||||||||||||||:::|||||
     305 ACTGAACCACCCTCAGTCTTCCTTGAGCGTCTCATGGAGGCATATAGGAG 354

     364 gPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAlaL 381
         |:::|||||||||||||||:::||||||:::|||||||||:::||||||:
     355 GTACACCCCTTTTGACCCCTCGTCAGAGGGGCAGAAAGCCGCTGTAGCCA 404

     381 euAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArg 397
         ::|||||||||||||||||||   ||||||:::|||||||||||||||||
     405 TGGCCTTCATTGGTCAGTCCGCTCCCGACATTAAGAAAAAGCTGCAAAGG 454

     398 LeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGl 414
         |||||||||||||||:::      |||:::|||||||||:::||||||||
     455 CTGGAGGGGCTCCAAGATCATACGCTCCAAGATTTAGTAAAAGAAGCAGA 504

     414 uLysValTyrTyrArgArgGluThrGluGluGluLysGluGlnArgLysG 431
         ||||||||||:::::|||||||||||||:::|||:::::::::||||::::
     505 GAAAGTCTATCATAAGAGGGAAACAGAAAAAGAGAGGCAGGAGAGAGAGA 554

     431 luLysGluArgGluGluArgGluGluArgArgAspArgArgGlnGlu 446
         ::|||||   ||||||||||||:::|||||||||||||||||||||
     555 AGAAAGAAATGGAGGAGAGGGAAAATAGACGGGATCGCCGTCAGGAG 601

seq_name: gb_gss:AZ708792

seq_documentation_block:
LOCUS       AZ708792      820 bp    DNA             GSS       24-JAN-2001
DEFINITION  RPCI-23-238P4.TJ RPCI-23 Mus musculus genomic clone RPCI-23-238P4,
            DNA sequence.
ACCESSION   AZ708792
VERSION     AZ708792.1  GI:12438716
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-238P4.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
```

```
           page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
           Plate: 238  row: P  column: 4
           Seq primer: SP6
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .820
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-238P4"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      236 a    186 c    247 g    151 t
ORIGIN


alignment_scores:
            Quality:  797.50            Length:    253
              Ratio:  3.853               Gaps:      5
 Percent Similarity:  81.818   Percent Identity: 65.217

alignment_block:
 US-09-171-553B-4 x AZ708792   ..

 Align seg 1/1  to: AZ708792  from: 1  to: 820

     280 ValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMe 296
         |||   |||   ||||||   ||||||:::|||   |||   ||||||
       1 GTCTTGGGGAACGATGGGACTCCCACCAACCTTCCTAACCTTATTGATGA 50

     296 tGlyPheProLeuThrArgProGlyTrpAspTyrAsnThrAlaGluGlyA 313
          :::||||||||||||||||||:::||||||||||||||||||||||||
      51 AGCTTTTCCTTTGACCCGCCCAAACTGGGACTACAATACAGCAGAAGGTA 100

     313 rgGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGly 329
         |||||   |||   :::||||||::::::|||||||||||||:::|||
     101 GGGAGCGTCTCCTGGTCTATCGCCGGACTCTAGTGGCGGGTCTCAAAGGG 150

     330 AlaSerArgArgProThrAsnLeuAlaLysValArgGluValMetGlnGl 346
         |||:::|||||||||      ||||||||||||||||||||:::|||||
     151 GCTGCACGACGGCCCTTCTTTCTGGCAAAGGTAAGAGAGGTCTTGCAGGG 200

     346 yProAsnGluProProSerValPheLeuGluArgLeuMetGluAlaPheA 363
         ||||   ||||||||||||||||||||||||||||||||||||||||:::|
     201 ACCAGCCGAGCCACCCTCGGTCTTTCTAGAACGGCTAATGGAGGCCTATA 250

     363 rgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerVal 379
         |||||:::|||||||||||||||:::||||||:::|||:::|||:::|||
     251 GGAGGTATACCCCGTTCGATCCTTCATCGGAGGGGCAGCAAGCAGCTGTT 300

     380 AlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGl 396
         |||:::|||||||||||||||||||||   ||||||:::||||||||||
     301 GCTATGGCTTTCATTGGCCAGTCAGCTCCCGATATAAAAAAGAAGCTGCA 350

     396 nArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluA 413
         |||||||||||||||||||:::      |||:::||||||||:::||||
     351 AAGACTCGAGGGCCTCCAAGATTATACTTTGCAGGATCTAGTGAAGGAAG 400

     413 laGluLysValTyrTyrArgArgGluThrGluGluGluLysGluGlnArg 429
         ||||||||||:::::::::||||||||||||:::|||||:::::|||
     401 CAGAGAAAGTGTTTCATAAGAGAGAGACGGAAGATGAAAAAAAGGAAAGA 450

     430 LysGluLysGluArgGluGluArgGluGluArgArgAspArgArgGlnGl 446
         ::::::||||||   |||:::|||||||:::||||||||||||||||||
     451 GAGAAAAAAGAGGCAGAAGACAGGGAGAATCGACGGGACCGTCGACAGGA 500

     446 uLysAsnLeuThrLysIleLeuAlaAlaValVal...GluGlyLysSerS 462
         |:::|||||||||:::||||||||||||||||||   ||||||:::
     501 AAGAAACCTGACAAGGATTCTGGCCGCAGTGGTAGGAGAAGGACAGTTAA 550

     462 erArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGln 478
           |||:::|||
     551 GGAGAGATAGG..................................... 561

     479 SerGlyAsnLeuGlyAsn.....................ArgThrPr 487
         ||||||||||||||||||                     :::||
     562 TCAGGGAACCTGGGCAACTGGCGAGAAGCCAGGATTGCAGACCCCAACCC 611

     487 o..LeuAspLysAspGlnCysAlaTyrCysLysGluLysGlyHisTrpAl 503
         |  |||:::|||||||||||||||||||||||||||:::|||||||||::
     612 CAGTTGGAAAAGGACCAGTGTGCGTATTGCAAAGAAAGGGGACATTGGGT 661

     503 aArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuG 520
         :|||:::|||||||||:::               ||||||||||||||:
     662 TCGGGAATGCCCAAAGAGAGA.........GCCAAGGTGCTGGCCCTCG 701

     520 luGluAsp 522
         :::::|||
     702 ATCAAGAT 709
```

OM of: US-09-171-553B-5 to: EST:* out_format : pfs

Date: Feb 24, 2002 8:31 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09171553/runat_22022002_122825_11137/app_query.fasta_1.3312
-DB=EST -QFMT=fastap -SUFFIX=feb22std.rst -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09171553_@CGN1_1_9008 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-171-553B-5
Query length: 1194
Database: EST:*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 3515.700000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_gss:AZ636636	-	1039.00	1567.71	3.7e-78	685	! AZ636636 1M0495O02R Mouse 10kb
gb_gss:BH090616	+	1011.50	1524.37	9.6e-76	791	! BH090616 RPCI-24-356L4.TJ RPCI
gb_est2:BI107577	+	1004.00	1512.85	4.2e-75	799	! BI107577 602894023F1 NCI_CGAP_I
gb_gss:BH056750	+	999.50	1505.20	1.1e-74	864	! BH056750 RPCI-24-94C16.TVB RPCI
gb_gss:AZ980516	-	998.00	1505.35	1.1e-74	682	! AZ980516 2M0257O06R Mouse 10kb
gb_est2:BI152889	+	993.00	1495.93	3.7e-74	813	! BI152889 602918048F1 NCI_CGAP_I
gb_gss:AZ627566	+	988.00	1490.01	7.9e-74	690	! AZ627566 1M0469L20F Mouse 10kb
gb_gss:AZ717731	+	981.00	1477.62	3.9e-73	817	! AZ717731 RPCI-24-155I1.TJ RPCI
gb_gss:AZ715956	+	972.50	1465.04	1.9e-72	789	! AZ715956 RPCI-24-155A1.TJ RPCI
gb_est2:BF784826	+	967.00	1454.49	7.5e-72	975	! BF784826 602111174F1 NCI_CGAP_K
gb_gss:BH106844	+	966.50	1455.24	6.8e-72	842	! BH106844 RPCI-24-352A17.TVB RPC
gb_gss:AZ715272	+	961.00	1447.99	1.7e-71	755	! AZ715272 RPCI-24-154N5.TJ RPCI
gb_gss:AZ827524	+	960.50	1448.01	1.7e-71	700	! AZ827524 2M0104E02F Mouse 10kb
gb_gss:AZ982997	-	952.00	1435.56	8.5e-71	668	! AZ982997 2M0264E02F Mouse 10kb
gb_htc:AK004796	+	950.00	1416.27	1.0e-69	3234	! AK004796 Mus musculus adult ma
gb_htc:AK004571	+	947.00	1411.52	1.9e-69	3293	! AK004571 Mus musculus adult ma
gb_gss:BH083773	+	941.00	1419.05	7.1e-70	653	! BH083773 RPCI-24-293G13.TV RPCI
gb_gss:AZ631391	+	939.50	1415.48	1.1e-69	740	! AZ631391 1M0485I06R Mouse 10kb
gb_gss:BH122548	+	938.50	1413.44	1.4e-69	778	! BH122548 RPCI-24-230J18.TJ RPCI
gb_gss:AZ353894	+	931.00	1404.04	4.8e-69	640	! AZ353894 1M0092P20R Mouse 10kb
gb_est1:AL134888	+	929.00	1401.12	7.0e-69	632	! AL134888 DKFZp762K0910_r1 762 (
gb_gss:AZ238859	-	922.00	1390.47	2.8e-68	632	! AZ238859 RPCI-23-87K22.TV RPCI
gb_gss:AZ866393	-	918.50	1383.73	6.5e-68	725	! AZ866393 2M0176E18R Mouse 10kb
gb_gss:AZ860071	+	917.00	1381.79	8.4e-68	701	! AZ860071 2M0165H23R Mouse 10kb
gb_gss:AZ974419	+	911.00	1372.64	2.7e-67	702	! AZ974419 2M0249C06F Mouse 10kb
gb_gss:BH042410	-	910.00	1370.88	3.4e-67	719	! BH042410 RPCI-24-301E24.TJ RPCI
gb_gss:AZ337037	+	908.00	1368.03	4.9e-67	705	! AZ337037 1M0067L03R Mouse 10kb
gb_gss:AZ618408	+	902.50	1359.47	1.5e-66	718	! AZ618408 1M0450C07F Mouse 10kb
gb_gss:BH092018	+	902.00	1358.25	1.7e-66	751	! BH092018 RPCI-24-356P6.TJ RPCI
gb_gss:AZ825369	+	900.50	1356.87	2.1e-66	688	! AZ825369 2M0100I04R Mouse 10kb
gb_gss:AZ962118	-	900.00	1357.25	2.0e-66	616	! AZ962118 2M0230N14R Mouse 10kb
gb_est2:X83313	-	898.00	1351.77	3.9e-66	780	! X83313 MMRS16REV Mouse cDNA Mus
gb_gss:AZ643009	+	891.00	1343.17	1.2e-65	639	! AZ643009 1M0506J17F Mouse 10kb
gb_gss:BH034484	+	891.00	1342.37	1.3e-65	691	! BH034484 RPCI-24-311D3.TJ RPCI
gb_est2:BI143830	+	891.00	1340.10	1.8e-65	861	! BI143830 602907110F1 NCI_CGAP_K
gb_gss:BH041228	+	890.00	1341.54	1.5e-65	646	! BH041228 RPCI-24-329L23.TJB RPC
gb_est2:BI408111	+	889.00	1335.65	3.1e-65	987	! BI408111 602918919F1 NCI_CGAP_I
gb_est2:BF162779	+	888.50	1338.18	2.3e-65	717	! BF162779 601769243F1 NCI_CGAP_I
gb_gss:BH074726	-	883.00	1329.57	6.8e-65	734	! BH074726 RPCI-24-331M11.TVB RPC
gb_gss:BH028043	+	883.00	1328.53	7.8e-65	812	! BH028043 RPCI-24-247H6.TV RPCI
gb_gss:AZ897994	-	881.00	1327.18	9.2e-65	689	! AZ897994 RPCI-24-173A19.TV
gb_gss:AZ403060	-	880.00	1325.96	1.1e-64	669	! AZ403060 1M0170L07R Mouse
gb_gss:AZ758880	+	876.50	1320.90	2.1e-64	652	! AZ758880 1M0551B09F Mouse
gb_est2:BI414234	+	876.50	1317.04	3.4e-64	948	! BI414234 602986684F1 NCI_C
gb_gss:AZ660560	+	872.00	1314.57	4.7e-64	620	! AZ660560 1M0538P10R Mouse

seq_name: gb_gss:AZ636636

seq_documentation_block:
LOCUS AZ636636 685 bp DNA GSS 13-DEC-2000
DEFINITION 1M0495O02R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0495O02 R, DNA sequence.
ACCESSION AZ636636
VERSION AZ636636.1 GI:11758826
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 685)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0495 row: O column: 02
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 685.
FEATURES Location/Qualifiers
source 1. .685
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0495O02"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 159 a 173 c 186 g 167 t
ORIGIN

alignment_scores:

```
           Quality: 1039.00            Length:    228
             Ratio:  4.833              Gaps:      0
 Percent Similarity:  94.298   Percent Identity: 84.211

alignment_block:
 US-09-171-553B-5 x AZ636636/rev   ..

 Align seg 1/1  to reverse of: AZ636636  from: 1  to: 685

     213 AsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPr 229
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     685 AATACACCACTTCTGCCGGTTCGAAAACCAGGGACCAATGACTATCGCCC 636

     229 oValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProT 246
         |||||||||||||||||||||||||||||||||   :::||||||||||||
     635 GGTACAAGACCTCCGGGAGGTTAACAAAAGGGTCCTGAACATTCACCCCA 586

     246 hrValProAsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsn 262
         |||||||||||||||||||||||||:|||||||:::|||||||||||||:::
     585 CAGTCCCGAACCCGTACAATTTATTAAGCTCTCTCCCACCTGAGAGAACA 536

     263 TrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHi 279
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     535 TGGTATACAGTTCTGGACTTAAAAGATGCCTTCTTTTGCCTGCGTTTGCA 486

     279 sProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyThrG 296
         ||||   ||||||   |||||||||||||||||||||||||||      |
     485 CCCTAAGAGTCAGCTCCTGTTTGCTTTTGAATGGAGGGACCCAGAGGGCG 436

     296 lyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsn 312
         ||:::||||||||||||||||||||||||||||||||||||||||||||||
     435 GACAGACTGGTCAACTAACTTGGACTAGGCTACCACAGGGGTTCAAAAAT 386

     313 SerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheAr 329
         |||||||||:::|||||||||||||||||||||||||||   |||||
     385 TCCCCCACCCTGTTTGACGAGGCCCTCCATCGGGATCTTGCGCCTTTTCG 336

     329 gIleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuL 346
         |   ::::::||||||:::|||||||||||||||||||||||||||||||:
     335 CGCTCGAAACCCTCAGCTTACCCTACTACAGTATGTGGATGATCTCTTGG 286

     346 euAlaGlyAlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeu 362
         :::::::::|||:::|||:::   |||   :::|||||||:::   ||||||
     285 TTGTGGCGGCCTCGAAGGAGCTGTGTCACCAGGGGACTGAGAGGCTCCTC 236

     363 LeuGluLeuSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIl 379
         |||||||||||||||||||||||||:::|||||||||||||||||||||||
     235 GCAGAACTGAGTGACTTGGGGTATCGAGTTTCGGCTAAAAAGGCACAAAT 186

     379 eCysArgArgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnA 396
         ||||:::   |||||||||||||||||||||||:::|||||||||||||:::|
     185 CTGTCAAACTGAGGTAACCTACCTGGGGTATACCCTCCGAGGGGGTAAAA 136

     396 rgTrpLeuThrGluAlaArgLysLysThrValValGlnIleProAlaPro 412
         ||||||||||||||||||||||||||||||||||:::::|||||:::|||
     135 GATGGCTCACAGAGGCCCGGAAGAAGACTGTTATGATGATCCCATCGCCA 86

     413 ThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysAr 429
         ||||||   :::|||||||||||||||||||||||||||||||||||||||
      85 ACTACCCCACGGCAGGTACGTGAGTTTCTGGGGACTGCTGGCTTTTGTAG 36

     429 gLeuTrpIleProGlyPheAlaThrLeuAlaAla 440
         |||||||||||||||||||||||||||||||||
      35 ACTCTGGATTCCAGGCTTTGCAACCCTAGCAGCA 2

seq_name: gb_gss:BH090616

seq_documentation_block:
LOCUS       BH090616      791 bp    DNA             GSS       18-JUL-2001
DEFINITION  RPCI-24-356L4.TJ RPCI-24 Mus musculus genomic clone RPCI-24-356L4,
            DNA sequence.
ACCESSION   BH090616
VERSION     BH090616.1  GI:14910521
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 356  row: L  column: 4
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-356L4"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      216 a    201 c    188 g    186 t
ORIGIN


alignment_scores:
           Quality: 1011.50            Length:    257
             Ratio:  4.341              Gaps:      2
 Percent Similarity:  90.661   Percent Identity: 73.541

alignment_block:
 US-09-171-553B-5 x BH090616   ..

 Align seg 1/1  to: BH090616  from: 1  to: 791

     118 ValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnVa 134
         ||||||:::|||||||||:::::|||||||||||||:::::      ::
      25 GTCTTAGCCCTGCAGCTTGAGGAAGAATACCGATTGCATGAA...GGCAT 71

     134 lLysProAspGln......AspIleGlnSerTrpLeuGluGlnPheProG 149
         :|||      :::      |||:::::::::|||||      ||||||:
      72 AAAGCAAACAGAGGTGCTAGACCTAAAAGACTGGTTGACTGCTTTCCCTA 121

     149 lnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPro 165
         ::|||||||||||||||:::|||||||||:::|||   :::||||||||||
     122 GAGCATGGGCAGAGACTGGAGGCATGGGAATGGCTGTCAGAGTTCCCCCT 171

     166 GlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTy 182
               :::   |||||:::::::|||||||||:::::::||||||||||||
     172 GTTTTTGTGGGACTGAAGACAGATGCAACCCCTATAGGAGTGCGCCAATA 221

     182 rProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArgL 199
         ||||:::|||||||||||||||:::|||||||   |||||||:::|||||||
     222 TCCAATGAGCCGTGAAGCAAGAGATGGAATCAGGCCACATATTCAGAGAC 271
```

```
    199 euIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPro 215
        ||:::|||   |||||||||||||||   ||||||||||||||||||||||
    272 TATTACAATTAGGTATTTTGGTACCGTGCCAGTCACCCTGGAATACTCCT 321

    216 LeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnAs 232
        ||||||||||||:::|||||||||||||:::||||||||||||||||||
    322 TTACTGCCGGTAAAGAAACCTGGCACGAGTGACTACCGACCAGTCCAAGA 371

    232 pLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProA 249
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    372 CCTGAGAGAGGTCAATAAGAGGGTCCAAGATATTCACCCAACTGTTCCTA 421

    249 snProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThr 265
        ||||||||||||||||||||:::||||||||||||||:::||||||||||
    422 ATCCCTATAACCTCCTCAGCTCACTCCCTCCAGAGCGAAAATGGTACACA 471

    266 ValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSe 282
        ||||||||||||||||||||||||||||||||||:::||||||||||::||
    472 GTACTTGATTTAAAAGATGCTTTCTTTTGCCTGAAGTTACATCCCTCCAG 521

    282 rGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrG 299
        |||||||:::|||||||||||||||||||||||||   ||||||:::   |
    522 CCAGCCCATATTCGCCTTTGAGTGGAGAGATCCTGACACTGGACAAATGG 571

    299 lyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThr 315
        ||||||||||||||||||||||||||||   ||||||||||||||||||
    572 GACAATTGACCTGGACACGGTTACCCCAGGAGTTCAAAAACTCCCCCACC 621

    316 IlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHi 332
        :::|||||||||||||||||||||||||||||||:::|||      :::::
    622 CTTTTTGATGAGGCTCTACATCGAGATTTAGCCTCCTTCCTAGCTGAAAA 671

    332 sProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyA 349
        :   |||||||||||||||||||||||||||||||||||||||||||||   :
    672 CTCCCAGGTAACCCTGCTCCAATATGTTGATGACTTGCTCCTTGCTGC.A 720

    349 laThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeu 365
        ::|||:::::::::|||   :::|||||||||   ||||||:::   |||
    721 CCACAGAGAGAGAATGTTGGCAGGGGACCAAGAGATTGCTAGTAGGACTG 770

    366 SerAspLeuGlyTyrArgAla 372
        ::::::|||||||||||||||
    771 GGTGAGTTGGGTTACCGGGCC 791

seq_name: gb_est2:BI107577

seq_documentation_block:
LOCUS       BI107577      799 bp    mRNA            EST       26-JUN-2001
DEFINITION  602894023F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5039044 5',
            mRNA sequence.
ACCESSION   BI107577
VERSION     BI107577.1  GI:14558470
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 799)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11107  row: n  column: 05
            High quality sequence stop: 743.

FEATURES             Location/Qualifiers
     source          1. .799
                     /organism="Mus musculus"
                     /strain="CZECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5039044"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      211 a    215 c    218 g    154 t      1 others
ORIGIN


alignment_scores:
            Quality: 1004.00             Length:    263
              Ratio:  4.236                Gaps:      5
 Percent Similarity:  90.114   Percent Identity: 75.285

alignment_block:
 US-09-171-553B-5 x BI107577   ..

 Align seg 1/1  to: BI107577  from: 1  to: 799

    507 ArgArgProValAlaTyrLeuSerLysLysLeuAspProValAlaSerGl 523
        ||||||||||||||   ||||||||||||||||||   ||||||||||
     20 CGGCGCCCGGTAGCCTAGTTATCAAAGAAACTGGACGCTGTGGCCAGTGG 69

    523 yTrpProValCysLeuLysAlaIleAlaAlaValAlaIleLeuValLysA 540
        |||||||   ||||||:::||||||||||||:::|||:::||||||:::|
     70 ATGGCCCTCCTGCCTGCGAGCGATAGCAGCCACGGCTGTGCTAGTAAGAG 119

    540 spAlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaProHis 556
        ||||||||||||||||||:::||||||||||:::|||:::::|||||||||
    120 ATGCTGACAAACTGACTATGGGCCAGAATGTTACTATAGTGGCCCCACAC 169

    557 AlaLeuGluAsnIleValArgGlnProProAspArgTrpMetThrAsnAl 573
        :::||||||:::|||:::||||||||||||||||||||||||||||||||
    170 TCTCTTGAGAGCATCATCAGGCAACCACCGGACCGCTGGATGACCAACGC 219

    573 aArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPheA 590
        ||||||||||||||||||||||||||||||||||||||||||||:::||||
    220 CCGAATGACGCACTACCAGAGCCTATTGCTGACAGAGCGAGTAAGTTTTG 269

    590 laProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAsp 606
        ||||||||||   ||||||||||:::|||||||||   |||:::|||
    270 CACCCCCAGCCATTCTCAACCCCGCCTCCTTACTACCT...GAGGCTGAC 316

    607 GluProValThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyVa 623
        |||      :::|||   |||::::::::|||   ||||||||||||::
    317 GAGGCCCCTGCACATAAGTGTGAAGAAATACTGGCAGAAGAGACTGGAAT 366

    623 lArgLysAspLeuThrAspIleProLeuThrGlyGluValLeuThrTrpP 640
        :|||   ||||||||||||   |||      |||   ::::::|||||||
    367 CCGGCCAGACCTCACAGACCAACCTTGGCCAGGG...GCGATGACTTGGT 413

    640 heThrAspGlySerSerTyrValValGluGlyLysArgMetAlaGlyAla 656
        ||||||||||||||||||:::||||||||||||||||||   |||||||||
    414 TCACGGACGGAAGCAGCTTTGTGGTAGAAGGTAAGCGGAAGGCTGGGGCT 463

    657 AlaValValAspGlyThr.ArgThrIleTrpAlaSerSerLeuProGluG 673
        ||||||||||||||||||    :::|||||||||||||||||||||||||
    464 GCAGTAGTGGATGGAACAGGCTGTCATATGGGCCAGCAGTCTGCCGGAGG 513

    673 lyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 689
        ||||||||||||||||||||||||||:::||||||   ||||||||||||
    514 GTACATCAGCTCAAAAAGCGGAACTAATCGCATTAATTCAAGCCTTAAGG 563

    690 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPh 706
```

```
      ||||||||||||:::::::::|||:::|||||||||||   ||||||||
    564 CTGGCAGAAGGAAGGGCTCTTAATGTCTATACCGACAGCCTGTACGCTTT 613

    706 eAlaThrAlaHisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuT 723
        ||||||||||||||||||||||||||||||:::::::||||||||||||
    614 TGCCACGGCTCATGTTCACGGAGCAATATACCGACACCGTGGACTGCTGA 663

    723 hrSerAlaGlyArgGluIleLysAsn.LysGluGluIleLeuSerLeuLe 739
        |||||   |||:::::::|||||||||    ||||||||||||||||||||
    664 CGTCTGACGGCAAAGATATCAAAAATACACGAAGAAATTCTCAGCTTATT 713

    739 uGluAla.LeuHisLeuProLysArgLeuAlaIleIleHisCysProGly 755
        ||||||| :::|||||||||:::|||:::|||||||||||||||||||||
    714 AGAAGCCTGTTCATTTGCCCCGTAGGGTGGCAATTATCCATTGCCCAGGA 763

    756 HisGlnLysAlaLysAspLeuIleSerArgGly 766
        ||||||||||:::         ::::::::|||
    764 CACCAGAAGGGACCTGCGCCCGTTGAAAAGGGA 796

seq_name: gb_gss:BH056750

seq_documentation_block:
LOCUS       BH056750      864 bp    DNA             GSS       18-JUL-2001
DEFINITION  RPCI-24-94C16.TVB RPCI-24 Mus musculus genomic clone RPCI-24-94C16,
            DNA sequence.
ACCESSION   BH056750
VERSION     BH056750.1  GI:14864050
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 864)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-94C16.TJ RPCI-24-94C16.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 94  row: C  column: 16
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .864
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-94C16"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      237 a    221 c    209 g    197 t
ORIGIN


alignment_scores:
```

```
            Quality:  999.50              Length:    295
              Ratio:  4.063                 Gaps:      2
Percent Similarity:  83.390   Percent Identity: 70.508

alignment_block:
US-09-171-553B-5 x BH056750   ..

Align seg 1/1  to: BH056750  from: 1  to: 864

    418 ValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGl 434
        ||||||||||||||||||||||||||||||||||||||||   ||||||||
      1 GTACGTGAGTTTCTGGGAACGGCAGGCTTTTGTAGACTCTGAATACCAGG 50

    434 yPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGlyG 451
        |   :::   |||||||||||||||||||||||||||||||||||||||
     51 GTCCACAATGTTAGCGGCACCTCTGTACCCACTCACTAAAGAAAAAGTCC 100

    451 lyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLys 467
        |||:::|||:::   ||||||||||||   |||::::::||||||
    101 CGTTCACCTGGACAGAAGAACACCAAAAAGACTTTAACAGCATAAAGGCT 150

    468 AlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPh 484
        |||||||||:::|||   ||||||||||||||   :::||||||||||||
    151 GCTCTGCTTGCAGCCCAGGCTTTGGCCCTGCCATATTTGACTAAGCCTTT 200

    484 eThrLeuTyrValAspGluArgLysGlyValAlaArgGlyValLeuThrG 501
        ||||:::|||||||||||||      ||||||||||||   |||||||||
    201 CACCATATACGTTGATGAGCCAGCTGGAGTAGCCCGCAGAGTCCTGACTC 250

    501 lnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLeu 517
        ||||||||||||||||||:::|||   ||||||||||||||||||||||||
    251 AGACTCTGGGGCCGTGGAAGAGGCTAGTGGCCTACCTGTCTAAAAAATTG 300

    518 AspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaAlaVa 534
        ||||||||||||||   ||||||   ||||||||||||||||||||||||
    301 GACCCAGTTGCCAGTGAATGGCCCTCCTGCCTAAAAGCCATCGCTGCAGT 350

    534 lAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIleT 551
        ||||:::|||||||||||||:::   |||||||||||||||||||:::::::
    351 AGCCCTACTTGTTAAAGATACTGGCAAGCTTACTCTAGGTCAGCATGTGA 400

    551 hrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAsp 567
        ::||||||:::||||||||||||:::::||||||   ||||||||||||
    401 GTGTAATCACTCCCCATGCGTTAAAGAGTATTGTATGACAGCCCCCTGAC 450

    568 ArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuTh 584
        :::||||||||||||||||||||||||||||||||||||||||||||::
    451 CATTGGATGACAAATGCCCGGATGACCCACTACCAGAGCCTGTTGCTAAA 500

    584 rGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeuL 601
        ::::|||||||||||||||   ||||||      ||||||:::|||||||
    501 TGACCGAGTGACTTTTGCCCTACCTGCCATCCACAATCCTACCACCCTTC 550

    601 euProGluGluThrAspGluProValThrHisAspCysHisGlnLeuLeu 617
        |||||:::      |||:::      :::|||:::|||   ::::::|||
    551 TCCCCAAGACGGATGACTCTACCTGGGTACACCAATGTGCTGATATCCTG 600

    618 IleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThrGl 634
        ||||||||||||   ||||||||||||||||||
    601 GATGAAGAAACTGGATCGAGAAAGGATCTGACTGAC.............. 636

    634 yGluValLeuThrTrp..................PheThrAspGlySerS 645
                |||||                     :::|||||||||||||
    637 .......CAAACTTGGCCAGGTGTGCCTGACAGGTACACAGATGGTAGCA 679

    645 erTyrValValGluGlyLysArgMetAlaGlyAlaAlaValValAspGly 661
        ||:::|||||||||||||||      ||||||   ||||||||||||
    680 GCTTTGTGGTGGAGGGGAAAATAAGAGCAGGAGAGGCAGTGGTTGACAGA 729

    662 ThrArgThrIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLy 678
        ::::::||||||||||||   ||||||||||||||||||||||||||||
    730 AAACATGTAATTTGGGCAAGTTAACTTCCAGAAGGAACTTCCGCTCAGAA 779
```

```
     678 sAlaGluLeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysS 695
         ||||||||||:::::||||||:::||||||   ||||||:::|||:::
     780 GGCTGAGCTCCTGACGCTAACTAAAGCTCTGCCGCTGGCAAAAGGTAGGC 828

     695 erIleAsnIleTyrThrAspSerArgTyrAlaPhe 706
           ||||||||||||||||||||:::|||||||||
     829 A.ATCAATATTTACACTGACAGCCACTATGCTTTT 863

seq_name: gb_gss:AZ980516

seq_documentation_block:
LOCUS       AZ980516      682 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0257O06R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0257O06 R, DNA sequence.
ACCESSION   AZ980516
VERSION     AZ980516.1  GI:13851743
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0257  row: O  column: 06
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 682.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0257O06"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      143 a    171 c    189 g    179 t
ORIGIN

alignment_scores:
           Quality:  998.00            Length:    227
             Ratio:  4.752               Gaps:      1
 Percent Similarity:  92.511   Percent Identity: 81.938

alignment_block:
 US-09-171-553B-5 x AZ980516/rev   ..

 Align seg 1/1  to reverse of: AZ980516  from: 1  to: 682

     179 ValArgGlnTyr.ProLeuSerArgGluAlaArgGluGlyIleTrpProH 195
         |||||||||||| |||:::|||:::|||||||||||||||||||   ||||
     681 GTAAGACAATACCCCCATGAGCAAGGAAGCTAGGGAGGGCATCCGGCCTC 632

     195 isValGlnArgLeuIleGlnGlnGlyIleLeuValProValGlnSerPro 211
         ||:::|||||||||:::::|||||||:::||||||      |||||||||
     631 ATATCCAGAGGTTGCTAGACCAAGGAGTTTTAGTGGCCTGTCAATCCCCC 582

     212 TrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrAr 228
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     581 TGGAATACACCACTTCTGCCGGTTCGAAAACCAGGGACCAATGACTATCG 532

     228 gProValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisP 245
         |||||||||||||||||||||||||||||||||   ||||||||||
     531 CCCGGTGCAAGACCTCCGGGAAGTTAACAAAAGGGTCCTGGACATTCACC 482

     245 roThrValProAsnProTyrAsnLeuLeuSerAlaLeuProProGluArg 261
         |||||||||||||||||||||||||||||||:::|||||||||||||||
     481 CCACAGTCCCGAACCCATACAATTTATTAAGCTCTCTCCCACCCGAGAGA 432

     262 AsnTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLe 278
         :::|||||||||||||||||||||||||||||||||||||||||||||||
     431 ACATGGTATACAGTCCTGGACTTAAAAGATGCCTTCTTTTGCCTGCGCTT 382

     278 uHisProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyT 295
         |||||||   ||||||   |||||||||||||||||||:::|||
     381 GCACCCTAAGAGTCAGCTCCTGTTTGCCTTTGAATGGAGGAACCCAGAGG 332

     295 hrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLys 311
           |||:::|||||||||||||||||||||||||||||||||||||||||||
     331 GCGGACAGACTGGTCAACTAACCTGGACTAGGCTACCACAGGGGTTCAAA 282

     312 AsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPh 328
         ||||||||||||:::|||||||||   ||||||||||||||||||   ||
     281 AATTCCCCCACCCTGTTTGACGAGGACCTCCATCGGGATCTCGCGCCTTT 232

     328 eArgIleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuL 345
         ||||   ::::::||||||:::|||||||||||||||||||||||||||||
     231 TCGCGCTCGAAACCCTCAGCTTACCCTACTACAGTATGTGGATGATCTCT 182

     345 euLeuAlaGlyAlaThrLysGlnAspCysLeuGluGlyThrLysAlaLeu 361
         ||:::|||:::|||:::|||:::   |||   :::||||||:::   |||
     181 TGGTCGCGGCAGCCTCGAAGGAGCTGTGTCACCAGGGAACTGAGAGGCTC 132

     362 LeuLeuGluLeuSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGl 378
         |||   ||||||||||||||||||||||:::|||||||||||||||||
     131 CTCACAGAACTGAGTGACTTGGGGTATCGAGTTTCGGCTAAAAAGGCACA 82

     378 nIleCysArgArgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyG 395
         |||||||::   |||||||||||||||||||||:::|||||||||||||:
      81 AATTTGTCAAACTGAGGTAACCTACCTGGGGTATACCCTCCGAGGGGGTA 32

     395 lnArgTrpLeuThrGluAlaArgLysLys 404
         ::|||||||||||||||||||||||||||
      31 AAAGATGGCTCACAGAAGCCCGGAAGAAG 3

seq_name: gb_est2:BI152889

seq_documentation_block:
LOCUS       BI152889      813 bp    mRNA            EST       05-JUL-2001
DEFINITION  602918048F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5068643 5',
            mRNA sequence.
```

```
ACCESSION   BI152889
VERSION     BI152889.1  GI:14612890
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11184  row: o  column: 12
            High quality sequence stop: 734.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Mus musculus"
                     /strain="CZECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5068643"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      225 a    218 c    205 g    165 t
ORIGIN


alignment_scores:
            Quality:  993.00            Length:    261
              Ratio:  4.208               Gaps:      4
 Percent Similarity:  90.421   Percent Identity: 76.628

alignment_block:
 US-09-171-553B-5 x BI152889   ..

 Align seg 1/1  to: BI152889  from: 1  to: 813

     150 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGl 166
         |||||||||||||||:::|||:::|||:::||||||   |||||||||
       2 GCGTGGGCAGAGACGGGAGGAGTGGGGATGGCAAAGTG.GTCCCCCCTGT 50

     166 nValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrP 183
          ||||||:::|||||:::::::|||||||||:::::::||||||||||||
      51 GGTGATTGAACTTAAGTCCGGGGCCACCCCTATAGGGGTCCGACAATATC 100

     183 roLeuSerArgGluAlaArgGluGlyIleTrpPro.HisValGlnArgLe 199
         ||:::||||||||||||:::|||||||||   ||| ::::::::::::||
     101 CCATGAGCAGAGAAGCTCAAGAGGGTATACGCCCTCCAAATCAACAAACT 150

     199 uIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProL 216
         |:::|||||||||||||||||||   :::||||||||||||||||||||
     151 GCTCCAACAAGGGATTTTGGTCCCATGCAAATCCCCTTGGAACACTCCTC 200

     216 euLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnAsp 232
         ||||||||||:::|||||||||||||:::|||||||||||||||||||||
     201 TACTTCCAGTAAAAAAACCAGGGACCAGGGACTATCGTCCAGTACAGGAC 250

     233 LeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAs 249
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 CTTAGAGAAGTCAACAAGAGAGTTCAGGACATACACCCCACGGTGCCAAA 300

     249 nProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrV 266
         ||||||||||||||||||:::|||||||||   |||:::|||||||||||
     301 TCCTTATAACCTCCTCAGCACCTTGCCACCTGGTCGGACATGGTACACAG 350

     266 alLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSer 282
         ||||||||||||||||||||||||||||||||||||||||||||:::|||
     351 TCCTGGATCTCAAAGACGCTTTTTTCTGTTTGAGGTTACACCCCAACAGC 400

     283 GlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGl 299
         |||||||||||||||||||||||||||||      :::|||::::::||
     401 CAGCCCTTGTTCGCTTTCGAATGGCGAGACTCCGAGAGTGGACAAGCCGG 450

     299 yGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrI 316
         ||||||||||||||||||||||||||||||||||||||||||||||||||:
     451 ACAGCTCACATGGACGAGGCTGCCTCAGGGATTCAAGAACTCGCCCACTT 500

     316 lePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGln.Hi 332
         ::|||||||||||||||||||||||||||||||   ||||||   ::: ::
     501 TGTTCGATGAAGCCCTACACCGAGATCTTGCTCTTTTCCGAGCCAAATAA 550

     332 sProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyA 349
         :|||||||||||||||||||||||||||||||||||||||||||||||:::
     551 CCCACAGGTGACTCTTCTGCAATATGTAGATGACCTGCTCCTAGCTGCAG 600

     349 laThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeu 365
           |||:::::::||||||      ||||||:::   ||||||   |||:::
     601 AAACACGCGAGGACTGTGAAATTGGGACCCAAAACCTCCTGGGCGAGGTT 650

     366 Ser..AspLeuGlyTyrArgAlaSerAla.LysLysAlaGlnIleCysAr 381
           :::|||||||||||||||||||||| ||||||||||||:::|||::
     651 AGGTAAGCCTGGGGTATCGGGCCTCTGCTTAAAAAGGCTCAGTTATGCCA 700

     381 gArgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpL 398
         :   ||||||||||||||||||||   ||||||   |||   |||   :
     701 GATAGAAGTGACCTACCTAGGATATGTCTTGAGAGATGGACCAAGGTGAG 750

     398 euThrGluAlaArgLysLysThrVal 406
         ::|||||||||||||   |||||||||
     751 TCACAGAAGCCAGAACAAAGACTGTA 776

seq_name: gb_gss:AZ627566

seq_documentation_block:
LOCUS       AZ627566      690 bp    DNA              GSS       13-DEC-2000
DEFINITION  1M0469L20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0469L20 F, DNA sequence.
ACCESSION   AZ627566
VERSION     AZ627566.1  GI:11749756
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0469  row: L  column: 20
            Seq primer: CGTTGTAAAACGACGGCCAGT
```

```
          Class: plasmid ends
          High quality sequence stop: 690.
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0469L20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      182 a    192 c    165 g    151 t
ORIGIN


alignment_scores:
            Quality:  988.00            Length:    228
              Ratio:  4.682               Gaps:      1
 Percent Similarity:  92.544   Percent Identity: 81.579

alignment_block:
 US-09-171-553B-5 x AZ627566   ..

 Align seg 1/1  to: AZ627566  from: 1  to: 690

     243 IleHisProThrValProAsnProTyrAsnLeuLeuSerAlaLeuProPr 259
         ||||||||||||||||||||||||||||||||||||||||:::|||||||||
       7 ATACACCCCACAGTGCCAAATCCTTATAACCTCCTCAGCACCTTGCCACC 56

     259 oGluArgAsnTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysL 276
         |   |||:::||||||||||||||||||||||||||||||||||||||||
      57 TGGTCGGACATGGTACACAGTCCTGGATCTCAAAGATGCTTTTTTCTGTT 106

     276 euArgLeuHisProThrSerGlnProLeuPheAlaPheGluTrpArgAsp 292
         ||||||||||||||:::|||||||||||||||||||||||||||||||||
     107 TGAGGTTACACCCCAACAGCCAGCCCTTGTTCGCTTTCGAATGGCGAGAC 156

     293 ProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGl 309
             :::|||::::::|||||||||||||||||||||||||||||||||||
     157 TCCGAGAGTGGACAAGCCGGACAGCTCACATGGACGAGGCTGCCTCAGGG 206

     309 yPheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuA 326
         |||||||||||||||||:::|||||||||||||||||||||||||||||
     207 ATTCAAGAACTCGCCCACTTTGTTCGATGAAGCCCTACACCGAGATCTTG 256

     326 laAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnTyrValAsp 342
         ||   ||||||   ::::::|||||||||||||||||||||||||||||||
     257 CTCTTTTCCGAGCCAATAACCCACAGGTGACTCTTCTGCAATATGTAGAT 306

     343 AspLeuLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGluGlyThrLy 359
         ||||||||||||||::   |||::::::||||||      |||||||::
     307 GATCTGCTCCTAGCTGCAGAAACACGCGAGGACTGTGAAATTGGGACCCA 356

     359 sAlaLeuLeuLeuGluLeuSerAspLeuGlyTyrArgAlaSerAlaLysL 376
         :   ||||||   ||||||:::   |||||||||||||||||||||||||
     357 AAACCTCCTGGGCGAGTTAGGTAAGCTGGGGTATCGGGCCTCTGCTAAAA 406

     376 ysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyrSerLeuArg 392
         ||||||||:::|||:::   ||||||||||||||||||||   ||||||
     407 AGGCTCAGTTATGCCAGATAGAAGTGACCTACCTAGGATATGTCTTGAGA 456

     393 GlyGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIl 409
         ||||||||||||||||||||||||||||||::::::|||:::||||||
     457 GATGGACAACGGTGGCTCACAGAAGCCAGAAAACAAGCTGTTATGCAGAT 506

     409 eProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaG 426
         ||||||||||||||||||:::|||||||||||||||||||||||||||||
     507 CCCGGCCCCAACCACTGCTCGCCAGGTAAGAGAGTTCCTGGGGACCGCCG 556

     426 lyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeu 442
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     557 GGTTTTGCAGACTCTGGATTCCCGGATTTGCCACACTGGCAGCTCCCTTA 606

     443 TyrProLeuThrLysGluLysGlyGly.PheSerTrpAlaProGluHisG 459
         |||||||||||||||||||||||||| |||:::|||:::   |||||||
     607 TATCCACTAACCAAAGAGAAAGGGGGAATTCACCTGGACCAGAGAACATC 656

     459 lnLysAlaPheAspAlaIleLysLysAlaLeu 469
         ||   ||||||::::::::::|||||||||||
     657 AGCTAGCCTTTGAAACTCTCAAAAAGGCACTG 688

seq_name: gb_gss:AZ717731

seq_documentation_block:
LOCUS       AZ717731      817 bp    DNA             GSS       24-JAN-2001
DEFINITION  RPCI-24-155I1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-155I1,
            DNA sequence.
ACCESSION   AZ717731
VERSION     AZ717731.1  GI:12456936
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 817)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-155I1.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 155  row: I  column: 1
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .817
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-155I1"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
```

```
                library was  cloned in the pTARBAC1 cloning vector at the
                BamH1 sites using MboI  partially digested male C57BL/6J
                DNA."
BASE COUNT      243 a    209 c    204 g    161 t
ORIGIN


alignment_scores:
            Quality:  981.00            Length:    276
              Ratio:  4.105               Gaps:      3
 Percent Similarity:  86.594   Percent Identity: 68.841

alignment_block:
 US-09-171-553B-5 x AZ717731   ..

 Align seg 1/1  to: AZ717731  from: 1  to: 817

     534 ValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIl 550
         ||||||:::|||:::||||||||||||||||||||:::||||||:::::
       3 GTAGCCCTGTTGATTAAAGATGCTGACAAATTGACAATGGGACAACAGGT 52

     550 eThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProA 567
         :||||||:::|||||||||||||||||||:::|||:::|||||||||||||
      53 GACTGTTGTGGCCCCTCATGCCTTAGAAAGTATCGCGCGTCAGCCACCTG 102

     567 spArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeu 583
         ||||||||||||||||||||||||||   ||||||||||||||||||||
     103 ACAGATGGATGACAAATGCCCGAATGCCACACTATCAGAGCCTGCTGCTA 152

     584 ThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLe 600
         :::||||||||||||||||||||||||||   ||||||||||||||||||
     153 AATGAGCGTGTAACCTTTGCGCCCCCTGCCATCCTCAACCCAGCTACCCT 202

     600 uLeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeuL 617
         |||||||      :::|||:::      :::|||:::|||   :::::::|
     203 TCTCCCTCTAACAAATGATTCCGTCCCAGTACATCAATGTATGGACATCC 252

     617 euIleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThr 633
         ||   ||||||||||||:::|||:::||||||||||||   |||
     253 TCGCTGAAGAAACTGGGACCAGAAGTGACCTGACTGACCAACCCTGGCCT 302

     634 GlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGluGl 650
         |||   :::   :::|||:::||||||||||||||||:::::::::|||||
     303 GGA...GCTCCCAGTTGGTACACGGACGGCAGCAGTTTCCTGATAGAGGG 349

     650 yLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpA 667
         |||||||   |||||||||||||||||||||||   ::::::|||||||
     350 AAAGCGAAAGGCTGGAGCTGCGGTGGTAGATGGGAAAAAGGTAATTTGGG 399

     667 laSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAla 683
         |||||:::|||||||||||||||||||||||||||||||||||||:::|||
     400 CAAGCGCTTTGCCTGAAGGAACGTCGGCACAAAAGGCTGAACTTATAGCG 449

     684 LeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrTh 700
         |||   ||||||||||||   |||:::||||||   ||||||||||||||
     450 CTTATTCAAGCCCTCCGAGAGGCTAAAGGTAAGATCATTAACATCTACAC 499

     700 rAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrL 717
         |||||||||||||||||||||||||||||:::|||||||||||||||||:
     500 TGACAGCCGCTATGCTTTTGCTACAGCACACATCCATGGGGCCATCTACA 549

     717 ysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGlu 733
         ::|||:::|||||||||||||||||||||||:::::|||||||||||||||
     550 GGCAGCAAGGGCTATTGACCTCGGCTGGTAAAGACATTAAAAACAAAGAA 599

     734 GluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIl 750
         |||||||||:::|||||||||||||:::|||   ||||||::::::|||||
     600 GAAATTCTGGCCCTGTTGGAAGCCATACATGCACCTAAGAAGGTAGCCAT 649

     750 eIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyA 767
         ||||||||||||||||||||:::::::::||||||:::::|||||||
     650 CATCCACTGCCCCGGCCACCAAAGAGGAGAAGACTTGGTGGCCAGGGGCA 699
```

```
     767 snGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeu 783
         ||:::|||||||||   ||||||||||||:::||||||::::::   :::
     700 ACCGAATGGCAGACTCAGTAGCAAAACAAAGTGCTCAAGGGGCCATGATC 749

     784 LeuProIleIleGluThrProLysAlaProGluProArg..........  796
         |||            |||   |||:::      |||:::
     750 TTA............ACTGAAAAAGGTGATCCGCCCAAAAGCACCTGAGA 787

     797 ....ArgGlnTyrThrLeuGluAspTrp 804
             |||   |||      |||   |||
     788 CGAGAGATATTACCATAAAGAGCTATGG 815

seq_name: gb_gss:AZ715956

seq_documentation_block:
LOCUS       AZ715956      789 bp    DNA             GSS       24-JAN-2001
DEFINITION  RPCI-24-155A1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-155A1,
            DNA sequence.
ACCESSION   AZ715956
VERSION     AZ715956.1  GI:12453181
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 789)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-155A1.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 155  row: A  column: 1
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .789
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-155A1"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      230 a    200 c    195 g    164 t
ORIGIN


alignment_scores:
            Quality:  972.50            Length:    253
              Ratio:  4.228               Gaps:      1
 Percent Similarity:  90.909   Percent Identity: 73.123

alignment_block:
 US-09-171-553B-5 x AZ715956   ..
```

```
Align seg 1/1  to: AZ715956  from: 1  to: 789

    527 CysLeuLysAlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLy 543
        ||||||||||||||||||||||||||:::|||:::|||||||||||||||
      5 TGTCTGAAAGCTATTGCAGCAGTAGCCCTGTTGATTAAAGATGCTGACAA 54

    543 sLeuThrLeuGlyGlnAsnIleThrValIleAlaProHisAlaLeuGluA 560
        |||||||:::|||||||:::::|||||:::|||||||||||||||||||:
     55 ATTGACAATGGGACAACAGGTGACTGTTGTGGCCCCTCATGCCTTAGAAA 104

    560 snIleValArgGlnProProAspArgTrpMetThrAsnAlaArgMetThr 576
        ::|||||||||||||||||||||||||||||||||||||||||||||||
    105 GTATCGTGCGTCAGCCACCTGACAGATGGATGACAAATGCCCGAATGCCA 154

    577 HisTyrGlnSerLeuLeuLeuThrGluArgValThrPheAlaProProAl 593
        ||||||||||||||||||||::::::   ||||||||||||||||||||
    155 CACTATCAGAGCCTGCTGCTAAATGATTGTGTAACCTTTGCGCCCCCTGC 204

    593 aAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProValT 610
        |   ||||||||||||||||||||||||      :::|||:::      :
    205 CATCCTCAACCCAGCTACCCTTCTCCCTCTAACAAATGATTCCGTCCCAG 254

    610 hrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAsp 626
        ::|||:::|||   :::::::|||   |||||||||||||:::|||:::|||
    255 TACATCAATGTATGGACATCCTCGCTGAAGAAACTGGGACCAGAAGTGAC 304

    627 LeuThrAspIleProLeuThrGlyGluValLeuThrTrpPheThrAspGl 643
        |||||||||    |||      |||    :::    :::|||:::|||||||||
    305 CTGACTGACCAACCCTGGCCTGGA...GCTCCCAGTTGGTACACGGACGG 351

    643 ySerSerTyrValValGluGlyLysArgMetAlaGlyAlaAlaValValA 660
        |||||||:::::::::|||||||||||   ||||||||||||||||||||
    352 CAGCAGTTTCCTGATAGAGGGAAAGCGAAAGGCTGGAGCTGCGGTGGTAG 401

    660 spGlyThrArgThrIleTrpAlaSerSerLeuProGluGlyThrSerAla 676
        |||||   ::::::|||||||||||||:::|||||||||||||||||||||
    402 ATGGGAAAAAGGTAATTTGGGCAAGCGCTTTGCCTGAAGGAACGTCGGCA 451

    677 GlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGl 693
        |||||||||||||||:::||||||   |||.|||||||||   |||:::||
    452 CAAAAGGCTGAACTTATAGCGCTTATTCAAGCCCTCCGAGAGGCTAAAGG 501

    693 yLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAlaH 710
        ||||   ||||||||||||||||||||||||||||||||||||||||||||
    502 TAAGATCATTAACATCTACACTGACAGCCGCTATGCTTTTGCTACAGCAC 551

    710 isValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGly 726
        ||:::|||||||||||||||:::|||:::|||||||||||||||||||||
    552 ACATCCATGGGGCCATCTACAGGCAGCAAGGGCTATTGACCTCGGCTGGT 601

    727 ArgGluIleLysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHi 743
        ::::::|||||||||||||||||||||||||:::|||||||||||||:::||
    602 AAAGACATTAAAAACAAAGAAGAAATTCTGGCCCTGTTGGAAGCCATACA 651

    743 sLeuProLysArgLeuAlaIleIleHisCysProGlyHisGlnLysAlaL 760
        |   ||||||:::::|||||||||||||||||||||||||||||:::::::
    652 TGCACCTAAGAAGGTAGCCATCATCCACTGCCCCGGGCACCAAAGAGGAG 701

    760 ysAspLeuIleSerArgGlyAsnGlnMetAlaAspArgValAlaLysGln 776
        ::||||||:::::::::|||:::      ||||||   ||||||||||||
    702 AAGACTTGGTGGGCCAGGGCCAACCGAATGCAGACTCAGTAGCAAAACAA 751

    777 AlaAlaGln 779
        :::|||||||
    752 GTTGCTCAA 760

seq_name: gb_est2:BF784826

seq_documentation_block:
LOCUS       BF784826      975 bp    mRNA            EST       12-JAN-2001
DEFINITION  602111174F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4239307
            5', mRNA sequence.
ACCESSION   BF784826
VERSION     BF784826.1  GI:12089862
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 975)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9852  row: c  column: 20
            High quality sequence stop: 701.
FEATURES             Location/Qualifiers
     source          1. .975
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4239307"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      257 a    281 c    252 g    185 t
ORIGIN


alignment_scores:
            Quality:  967.00              Length:    349
              Ratio:  3.748                 Gaps:      8
 Percent Similarity:  73.926   Percent Identity: 60.172

alignment_block:
 US-09-171-553B-5 x BF784826   ..

 Align seg 1/1  to: BF784826  from: 1  to: 975

    157 MetGlyLeuAlaLysGlnValProProGlnValIleGlnLeuLysAlaSe 173
        :::|||:::||||||   |||||||||   |||||||:::||||||::::
      3 GTGGGGATGGCAAAACTGGTCCCCCCCCGTGGTGATTGAACTTAAGTCCGG 52

    173 rAlaThrProValSerValArgGlnTyrProLeuSerArgGluAlaArgG 190
        :|||||||||:::::|||||||||||||:::|||||||||||||:::|
     53 GGCCACCCCTATAGGGGTCCGACAATATCCCATGAGCAGAGAAGCTCAAG 102

    190 luGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyIleLeuVal 206
        ||||||||   |||:::::::::::::|||:::|||||||||||||||||
    103 AGGGTATACGCCCCCAAATTAACAAACTGCTCCAACAAGGGATTTTGGTC 152

    207 ProValGlnSerProTrpAsnThrProLeuLeuProValArgLysProGl 223
        |||   :::|||||||||||||||||||||||||||||||||:::|||
    153 CCATGCAAATCCCCTTGGAACACTCCTCTACTTCCAGTAAAAAAAACCAG 202

    223 yThrAsnAspTyrArgProValGlnAspLeuArgGluValAsnLysArgV 240
           :::|||||||||||||||||||||||||||||||||||||||||||||
    203 GGACCAGGACTACCGTCCAGTACAAGACCTTAGAGAAGTCAACAAGAGAG 252

    240 alGlnAspIleHisProThrValProAsnProTyrAsnLeuLeuSerAla 256
        |||||||||||||||||||||||||||||||||||||||||||||||:::
    253 TTCAGGACATACACCCCACGGTGCCAAATCCTTATAACCTCCTCAGCACC 302

    257 LeuProProGluArgAsnTrpTyrThrValLeuAspLeuLysAspAlaPh 273
        |||||||||   |||:::|||||||||||||||||||||||||||||||
```

```
     303 TTGCCACCTGGTCGGACATGGTACACAGTCCTGGATCTCAAAGACGCTTT 352

     273 ePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGluT 290
         ||||||||||||||||||||||:::||||||||||||||||||||||||||
     353 TTTCTGTTTGAGGTTACACCCCAACAGCCAGCCCTTGTTCGCTTTCGAAT 402

     290 rpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeu 306
         ||||||||      :::|||:::::||||||||||||||||||||||||
     403 GGCGAGACTCCGAGAGTGGACAAGCCGGACAGCTCACATGGACGAGGCTG 452

     307 ProGlnGlyPheLysAsnSerProThrIlePheAspGluAlaLeuHisAr 323
         ||||||||||||||||||||||||||:::||||||||||||||||||||
     453 CCTCAGGGATTCAAGAACTCGCCCACTTTGTTCGATGAAGCCCTACACCG 502

     323 gAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnT 340
         ||||||||||   ||||||   ::::::||||||||||||||||||||
     503 AGATCTTGCTCTTTTCCGAGCCAATAACCCACAGGTGACTCTTCTGCAAT 552

     340 yrValAspAspLeuLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGlu 356
         |||||||||||||||||||||||:::   |||:::::|||||
     553 ATGTAGATGACCTGCTCCTAGCTGCAGAAACACGCGAGGACTGTGAAATT 602

     357 GlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGlyTyrArgAlaSe 373
         ||||||:::            :::         ||||||||||||||||
     603 GGGACCCAAAACTCCTGGGCGAGTTAGGTAAAGCTGGGGTATCGGGCCTC 652

     373 rAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyrS 390
         ||||   |||||||||:::|||:::   ||||||||||||||||||||
     653 TGCTAA.AAGGCTCAGTTATGCCAGATAGAAGTGACCTACCTAGGATATG 701

     390 erLeuArgGly.GlyGlnArgTrpLeuThrGluAlaArgLysLysThrVa 406
           ||||||::: |||   ||||||      :::   ::::::||||||
     702 TCTTGCGAGCATGGGCCACGGTGG......TCCCAGAAGCAGAAAACAGG 745

     406 lValGlnIleProAlaProThrThrAlaLysGlnValArgGluPheLeuG 423
          |||      |||   ||||||   :::            |||   ||||
     746 TGTTGCAGATCCCGAGCCCACAATGTCGCCAGTA......GAAGTCCTGG 789

     423 lyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAla 439
         ||      |||   |||   |||
     790 GGCCGCCGGGT...TGCGACCTG..........................  809

     440 AlaProLeuTyrProLeuThrLysGluLysGlyGlyPheSerTrpAlaPr 456
            |||:::   |||                          |||:::||
     810 ATCCCGATTTGCCCA.........................TGGGGTCC 832

     456 oGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAlaP 473
         |          |||      ::::::      ||||::
     833 TTGTTCCCTACCAAGAAGGGATCCGCGCCAAGAACACTGCTTGAACCAAA 882

     473 roAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 489
           :::         |||   ||||||:::|||
     883 GCGGGGCGCGGGGCCCA...GTAACCTCCCCT.................. 911

     490 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGly 504
         :::::::::|||      ||||||      ||||||      |||
     912 CGGGAAAGAGGGGGCCACCGCGGACGAGCGACACAAAAAAGGGGC 956

seq_name: gb_gss:BH106844

seq_documentation_block:
LOCUS       BH106844      842 bp    DNA             GSS       19-JUL-2001
DEFINITION  RPCI-24-352A17.TVB RPCI-24 Mus musculus genomic clone
            RPCI-24-352A17, DNA sequence.
ACCESSION   BH106844
VERSION     BH106844.1  GI:14937281
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 842)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-352A17.TJB
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 352  row: A  column: 17
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .842
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-352A17"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      221 a    228 c    200 g    193 t
ORIGIN


alignment_scores:
           Quality:  966.50              Length:    267
             Ratio:  4.221                 Gaps:      1
 Percent Similarity:  85.768   Percent Identity: 70.412

alignment_block:
 US-09-171-553B-5 x BH106844   ..

 Align seg 1/1  to: BH106844  from: 1  to: 842

     373 SerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTy 389
         ::::::|||||||||||||:::|||:::   ::::::|||||||||||||
      18 ACCACTAAGAAAGCTCAGCTATGCCATTTGGACTTGACCTACCTAGGATA 67

     389 rSerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysLysThrV 406
         ||||||||||   |||::::::|||:::|||||||||||||||||||||
      68 CTCCCTCCGAGAAGGAAAGCAGTGGTTCACCGAAGCTCGAAAGAAGACTG 117

     406 alValGlnIleProAlaProThrThrAlaLysGlnValArgGluPheLeu 422
         ||:::||||||||:::||||||||      ||||||||||||||||||
     118 TGATGCAGATCCCAACCCCCACCACTCTGTGACAAGTACGTGAGTTTCTG 167

     423 GlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAl 439
         ||||||||||||||||||||||||   |||||||||   :::   |||||
     168 GGAACGGCAGGCTTTTGTAGACTCTGAATACCAGGGTCCACAATGTTGGC 217

     439 aAlaProLeuTyrProLeuThrLysGluLysGlyGlyPheSerTrpAlaP 456
         ||||||||||||||||||||||||||||||      |||:::|||:::
     218 GGCACCTCTGTACCCACTCACTAAAGAAAAAGTCCCGTTCACCTGGACAG 267

     456 roGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAla 472
         ||||||||||||||||||:::::::||||||   |||||||||:::|||
     268 AGGAACACCAAAAAGCCTTTAACAGCATAAAGGCTGCTCTGCTTGCAGCC 317

     473 ProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAs 489
```

```
                  ||||||||||||||||||:::|||||||||||||||||:::||||||||
     318 CAGGCTTTGGCCCTGCCAGATTTGACTAAGCCTTTCACCATATACGTTGA 367

     489 pGluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProT 506
         ||||     ||||||||||||   ||||||||||||||||||||||||||
     368 TGAGCCAGCTGGAGTAGCCCGCAGAGTCCTGACTCAGACTCTGGGGCCCT 417

     506 rpArgArgProValAlaTyrLeuSerLysLysLeuAspProValAlaSer 522
         ||:::||||||||||||||||||||||||||||||||||||||||||||
     418 GGAAGAGGCCAGTGGCCTACCTGTCTAAAAAATTGGACCCAGTTGCCAGT 467

     523 GlyTrpProValCysLeuLysAlaIleAlaAlaValAlaIleLeuValLy 539
             ||||||   ||||||||||||||||||:::||||||:::||||||||
     468 GAATGGCCCTCCTGCCTAAAAGCCATCGCTGTAGTAGCCCTCCTTGTTAA 517

     539 sAspAlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaProH 556
         ||||:::   |||:::||||||||||||||:::::::::::|||:::||||
     518 AGATACTGGCAAGTTTACTCTAGGTCAGCATGTGAGTTTAATCACTCCCC 567

     556 isAlaLeuGluAsnIleValArgGlnProProAspArgTrpMetThrAsn 572
         ||||||||:::::::|||||   ||||||||||||:::||||||||||||
     568 ATGCGTTAAAGAGTATTGTATGACAGCCCCCTGACCATTGGATGACAAAT 617

     573 AlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPh 589
         ||||||||||||||||||||||||||||||||:::::|||||||||||||
     618 GCCCGGATGACCCACTACCAGAGCCTGTTGCTAAATGACCGAGTGACTTT 667

     589 eAlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrA 606
         ||||   ||||||      ||||||:::||||||||||||:::
     668 TGCCCTACCTGCCATCCACAATCCTACCACCCTCCTTCCCAAGACGGATG 717

     606 spGluProValThrHisAspCysHisGlnLeuLeuIleGluGluThrGly 622
           :::      :::|||:::|||   :::::::|||   ||||||||||||
     718 CCTCTACCTGGGTACACCAATGTGCTGATATTCTGGCTGAAGAAACTGGA 767

     623 ValArgLysAspLeuThrAspIleProLeuThrGlyGluValLeuThrTr 639
         ||||||||||||||||||                          |||||
     768 TCGAGAAAGGATCTGACTGAC.....................CGAACTTG 796

     639 p 639
         |
     797 G 797
```

seq_name: gb_gss:AZ715272

seq_documentation_block:

```
LOCUS       AZ715272      755 bp    DNA             GSS       24-JAN-2001
DEFINITION  RPCI-24-154N5.TJ RPCI-24 Mus musculus genomic clone RPCI-24-154N5,
            DNA sequence.
ACCESSION   AZ715272
VERSION     AZ715272.1  GI:12451823
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 755)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-154N5.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 154  row: N  column: 5
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .755
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-154N5"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      202 a    194 c    190 g    169 t
ORIGIN
```

alignment_scores:

```
            Quality:  961.00              Length:    253
              Ratio:  4.271                 Gaps:      3
 Percent Similarity:  88.933   Percent Identity: 75.099
```

alignment_block:

US-09-171-553B-5 x AZ715272 ..

Align seg 1/1 to: AZ715272 from: 1 to: 755

```
     273 PhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGl 289
         ||||||||||||         |||:::||||||||||||||||||||||||
       5 TTTTTCTGTTTGGGT...TCACCCAACAGCCAGCCCTTGTTCGCTTTCGA 51

     289 uTrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgL 306
         ||||   |||       :::|||::::::|||||||||||||||||||||
      52 ATGGTGAGACTCCGAGAGTGGACAAGCCGGACAGCTCACATGGACGAGGC 101

     306 euProGlnGlyPheLysAsnSerProThrIlePheAspGluAlaLeuHis 322
         ||||||||||||||||||||||||||||||:::|||||||||||||||||
     102 TGCCTCAGGGATTCAAGAACTCGCCCACTTTGTTCGATGAAGCCCTACAC 151

     323 ArgAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGl 339
         ||||||||||||   |||:::   ::::::|||||||||   ||||||||
     152 CGAGATCTTGCTCTTTTCCAAGCCAATAACCCACAGGTGATTCTTCTGCA 201

     339 nTyrValAspAspLeuLeuLeuAlaGlyAlaThrLysGlnAspCysLeuG 356
         ||||||||||||||||||||||||||:::   |||::::::||||||
     202 ATATGTAGATGACCTGCTCCTAGCTGCAGAAACACGTGAGGACTGTGAAA 251

     356 luGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGlyTyrArgAla 372
           ||||||:::   ||||||   ||||||:::   ||||||||||||||||
     252 TTGGGACCCAAAACCTCCTGGGCGAGTTAGGTAAGCTGGGGTATCGGGCC 301

     373 SerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTy 389
         ||||||||||||||||||:::|||:::   :::||||||||||||||||||
     302 TCTGCTAAAAAGGCTCAGTTATGCCAGATAAAAGTGACCTACCTAGGATA 351

     389 rSerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysLysThrV 406
         |   ||||||   ||||||||||||||||||:::::|||||:::::|
     352 TGTCTTGAGAGATGGACAACGGTGGCTCACAAAAACCAGAAAACAAGCTG 401

     406 alValGlnIleProAlaProThrThrAlaLysGlnValArgGluPheLeu 422
         ||:::||||||||:::||||||||||||:::||||||||||||||||||||
     402 TTATGCAGATCCCGACCCCAACCACTGCTCGCCAGGTAAGAGAGTTCCTG 451

     423 GlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAl 439
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     452 GGGACCGCCGGGTTTTGCAGACTCTGGATTCCCGGATTTGCCACACTAGC 501
```

```
      439 aAlaProLeuTyrProLeuThrLysGluLysGlyGlyPheSerTrpAlaP 456
          |||||||||||||||||||||||||||||||||   |||:::|||
      502 AGCTCCCTTGTATCCACTAACCAAAGAGAAAGGGGAATTCACCTGGATAG 550

      456 roGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAla 472
            |||||||||   ||||||::::::::::|||||||||||||||:::|||
      551 A.GAACATCAGCTAGCCTTTGAAACTCTCAAAAAGGCACTGCTGCAGGCT 600

      473 ProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAs 489
          |||||||||||||||||||||:::::::||||||||||||||||||:::||
      601 CCGGCATTGGCCCTGCCAGATTTAAACAAACCTTTCACCCTATACATTGA 650

      489 pGluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProT 506
          |||||||:::|||||||||||||||   |||||||||:::||||||||||
      651 TGAAAGAAATGGAGTGGCAAGGGGAGGCCTTACCCAGGTTTTGGGACCAT 700

      506 rpArgArgProValAlaTyr.LeuSerLysLysLeuAspProValAlaSe 522
          ||:::|||||||||||||||       |||:::   |||   ||||||||
      701 GGAAGCGCCCGGTAGCCTACTTAT...AAAGAAACTGACGCTGTGGCCAG 747

      522 rGlyTrp 524
          |||||||
      748 TGGATGG 754

seq_name: gb_gss:AZ827524

seq_documentation_block:
LOCUS       AZ827524      700 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0104E02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0104E02 F, DNA sequence.
ACCESSION   AZ827524
VERSION     AZ827524.1  GI:12997432
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0104  row: E  column: 02
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 700.
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0104E02"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      178 a    182 c    175 g    165 t
ORIGIN


alignment_scores:
           Quality:  960.50              Length:    233
             Ratio:  4.488                 Gaps:      1
 Percent Similarity:  91.845   Percent Identity: 76.824

alignment_block:
 US-09-171-553B-5 x AZ827524   ..

 Align seg 1/1  to: AZ827524  from: 1  to: 700

      416 LysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIl 432
          :::||||||||||||||||||||||||||||||||||||||||||||||||
        5 CGGCAGGTACGTGAGTTTCTGGGGACTGCTGGCTTTTGTAGACTCTGGAT 54

      432 eProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluL 449
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       55 TCCAGGCTTTGCAACCCTAGCAGCACCTCTATATCCTTTGACTAAGGAAG 104

      449 ysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIle 465
                |||:::|||       |||||||||:::|||||:::|||||||
      105 GGGTTCCTTTCGAGTGGAAAGAAGAACACCAGAGAGCTTTTGAGGCTATC 154

      466 LysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLy 482
          |||:::::::|||:::::::|||||||||||||||||||||||:::|||||
      155 AAGTCGTCTCTAATGACTGCCCCCGCGCTAGCATTACCAGACTTGACTAA 204

      482 sProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyValL 499
          |||||||:::|||||||||||||||||||   |||||||||||||||||||
      205 GCCTTTCGTCCTATATGTGGACGAGAGAGCGGGTGTAGCCAGGGGAGTGT 254

      499 euThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLys 515
          ||||||||:::|||||||||||||:::|||||||||||||||||||||||
      255 TGACACAAGCACTGGGACCCTGGAAGAGACCTGTAGCCTATTTGTCAAAA 304

      516 LysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAl 532
          ||||||||||||||||||||||||||||||:::|||||||||||||||||
      305 AAATTAGATCCTGTTGCTAGTGGATGGCCCACATGTCTGAAAGCTATTGC 354

      532 aAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnA 549
          ||||||||||:::|||:::|||||||||||||||||||||:::|||||:
      355 AGCAGTAGCCCTGTTGATCAAAGATGCTGACAAATTGACAATGGGACAAC 404

      549 snIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPro 565
          :::::||||||:::|||||||||||||||||||:::|||||||||||||||
      405 AGGTGACTGTTGTGGCCCCTCATGCCTTAGAAAGTATCGTGCGGCAGCCA 454

      566 ProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLe 582
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      455 CCTGACAGATGGATGACAAATGCCCGAATGACACACTATCAGAGCCTGCT 504

      582 uLeuThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaT 599
          ||||:::|||||||||||||||||||||||||   ||||||||||||||
      505 GCTAAATGAGCGTGTAACCTTTGCGCCCCCTGCCATCCTCAACCCAGCTA 554

      599 hrLeuLeuProGluGluThrAspGluProValThrHisAspCysHisGln 615
          ||||||||||||      :::|||:::      :::|||:::|||   :::
```

```
     555 CCCTTCTCCCTCTAACAAATGATTCCGTCCCAGTACATCAATGTATGGAC 604

     616 LeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleProLe 632
         :::|||   ||||||||||||:::|||:::||||||||||||   |||
     605 ATCCTTGCTGAAGAAACTGGGACCAGAAGTGACCTGACTGACCAACCCTG 654

     632 uThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValVal 648
             |||   :::   :::|||:::|||||||||   |||:::::::::
     655 GCCTGGA...GCTCCCAGTTGGTACACGGACGGCAGGAGTTTCCTGATA 700

seq_name: gb_gss:AZ982997

seq_documentation_block:
LOCUS       AZ982997      668 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0264E02F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0264E02 F, DNA sequence.
ACCESSION   AZ982997
VERSION     AZ982997.1  GI:13854224
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0264  row: E  column: 02
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 668.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0264E02"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      148 a    163 c    188 g    169 t

ORIGIN


alignment_scores:
            Quality:  952.00              Length:    212
              Ratio:  4.832                 Gaps:      0
 Percent Similarity:  92.925   Percent Identity: 83.019

alignment_block:
 US-09-171-553B-5 x AZ982997/rev   ..

 Align seq 1/1  to reverse of: AZ982997  from: 1  to: 668

     202 GlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLeuPr 218
         ||||||:::||||||      ||||||||||||||||||||||||||||||
     665 CAAGGAGTTTTAGTGGCCTGTCAGTCCCCCTGGAATACACCACTTCTGCC 616

     218 oValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgG 235
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     615 AGTTCGAAAACCAGGGACCAATGACTATCGCCCGGTGCAAGACCTCCGGG 566

     235 luValAsnLysArgValGlnAspIleHisProThrValProAsnProTyr 251
         ||||||||||||||||   ||||||||||||||||||||||||||||||
     565 AAGTTAACAAAAGGGTCCTGGACATTCACCCCACAGTCCCGAACCCGTAC 516

     252 AsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLeuAs 268
         ||||||||||||:::||||||||||||:::||||||||||||||||||
     515 AATTTATTAAGCTCTCTCCCACCCGAGAGAACATGGTATACAGTCCTGGA 466

     268 pLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProL 285
         ||||||||||||||||||||||||||||||||||||   ||||||   |
     465 CTTAAAAGATGCCTTCTTTTGCCTGCGCTTGCACCCTAAGAGTCAGCTCC 416

     285 euPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGlnLeu 301
         ||||||||||||||||||||||||||         :::|||||||||||||
     415 TGTTTGCCTTTGAATGGAGGGACCCAGAGGGCAGACAGACTGGTCAACTA 366

     302 ThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePheAs 318
         ||||||||||||||||||||||||||||||||||||||||||:::||||||
     365 ACCTGGACTAGGCTACCACAGGGGTTCAAAAATTCCCCCACCCTGTTTGA 316

     318 pGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnV 335
         ||||||||||||||||||||||||   ||||||   ::::::||||||:
     315 CGAGGCCCTCCATCGGGATCTCGCGCCTTTTCGCGCTCGAAACCCTCAGC 266

     335 alThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAlaThrLys 351
         ::||||||||||||||||||||||||||||||:::|||:::|||:::|||
     265 TTACCCTACTACAGTATGTGGATGATCTCTTGGTCGCGGCGGCCTCGAAG 216

     352 GlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLe 368
         :::   |||   :::||||||:::   ||||||   ||||||||||||||
     215 GAGCTGTGTCACCAGGGAACTGAGAGGCTCCTCACAGAACTGAGTGACTT 166

     368 uGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValT 385
         ||||||||||:::||||||||||||||||||||||||:::   |||||||
     165 GGGGTATCGAGTTTCGGCTAAAAAGGCACAAATTTGTCAAACTGAGGTAA 116

     385 hrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrGluAla 401
         ||||||||||||||:::||||||||||||:::||||||||||||||||||
     115 CCTACCTGGGGTATACCCTCCGAGGGGGTAAAAGATGGCTCACAGAAGCC 66

     402 ArgLysLysThrValValGlnIleProAlaProThr 413
         ||||||||||||||:::::::||||||:::||||||
      65 CGGAAGAAGACTGTTATGATGATCCCATCGCCAACG 30

seq_name: gb_htc:AK004796

seq_documentation_block:
LOCUS       AK004796     3234 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male lung cDNA, RIKEN full-length enriched
            library, clone:1200015M12, full insert sequence.
ACCESSION   AK004796
VERSION     AK004796.1  GI:12836246
```

KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male lung cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:1200015M12.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3234)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 3234)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 3234)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (bases 1 to 3234)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 3234)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

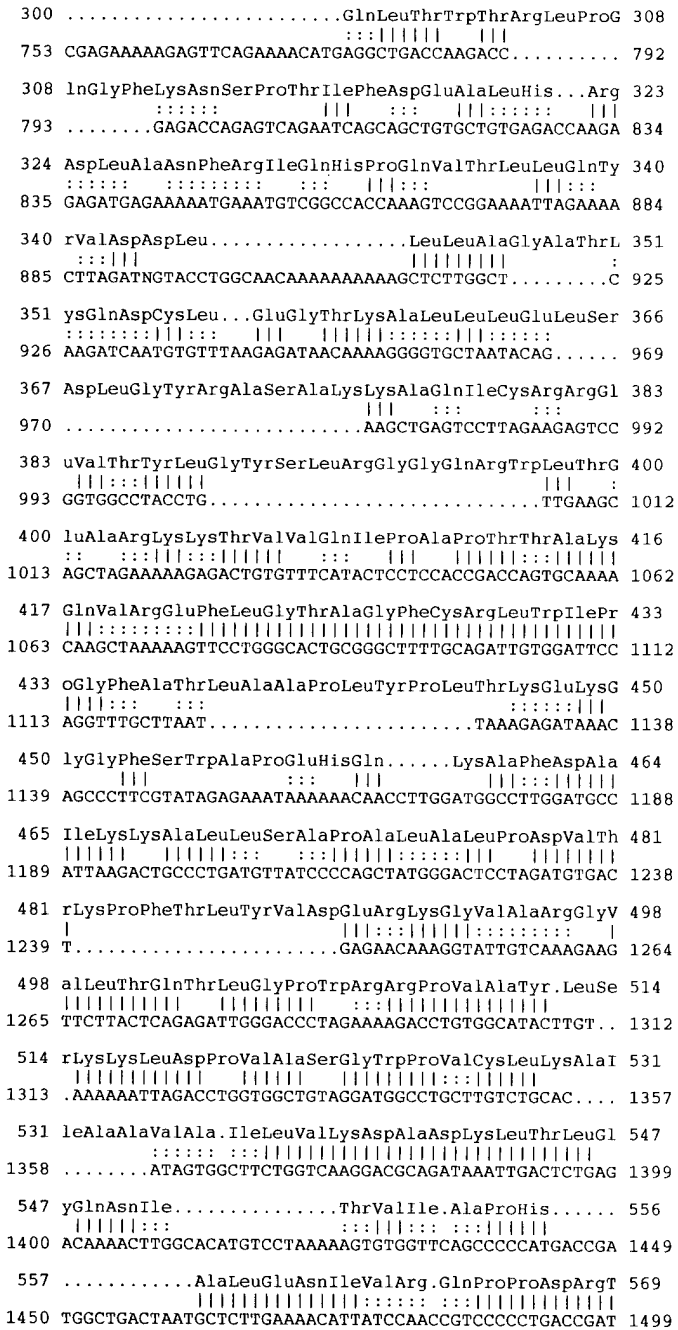
FEATURES Location/Qualifiers
source 1. .3234
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1904372"
/db_xref="MGD:MGI:1918989"
/clone="1200015M12"
/sex="male"
/tissue_type="lung"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
BASE COUNT 1005 a 688 c 749 g 791 t 1 others
ORIGIN

alignment_scores:
Quality: 950.00 Length: 1090
Ratio: 1.715 Gaps: 58
Percent Similarity: 50.826 Percent Identity: 33.028

alignment_block:
US-09-171-553B-5 x AK004796 ..

Align seg 1/1 to: AK004796 from: 1 to: 3234

```
    173 SerAlaThrProValSerValArgGlnTyrProLeuSerArgGluAlaAr 189
        |||:::||||||   |||:::   :::::||||||   |||:::
    243 AGCTGCACTCCCTTCTCTGCTCTAAAGTTCCCTCTTCTCAGAAGG..... 287

    189 gGluGlyIleTrpProHisValGlnArgLeu..................... 199
                  |||   |||   :::   |||
    288 ..........TGGCACCACCCTGAGTTGCTGGCAGTGAGTCTGTTCCAAG 327

    200 .....................IleGlnGlnGlyIleLeuValProValGln 209
                                  :::   |||:::|||      |||:::
    328 TTCCAGTTAGGGAGGCATCTGCCCACTTGGGGCTTCTGTCCAAGGTAAGG 377

    210 Ser.Pro.........................TrpAsnThrProLeuLeuP 218
        ||| |||                        |||::::::   |||
    378 AGCACCTGTGAGTCTAACTGCCAGGCTCTGATGGGGGTCTCATCTCTGCG 427

    218 roValArgLysProGlyThrAsnAspTyr......ArgProValGlnAsp 232
          :::||||||         :::|||         |||   :::|||
    428 GGACTAGAAAGTCCCAACAATCTGACCAAGGTAACAGGAAGTTACAAAAG 477

    233 LeuArgGlu................................ValAsnLy 238
          ||||||                                |||:::||
    478 CAAAGAGAAGGACAGGGTAAGTTGGTAAGAAAGTTAGTGTAAGTAGATAA 527

    238 sArgValGlnAspIleHisProThrValProAsnProTyrAsnLeuLeuS 255
        |:::|||||
    528 GAAAGTGCAG..................................... 537

    255 erAlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeuLysAsp 271
                  :::||||||         :::         |||:::
    538 ...........AGGAGGAGAAACAAAAGAAGGATAGAGGTCTGCAAAGAA 576

    272 AlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPh 288
        :::                        :::|||:::   :::|||:::
    577 GTG.......................AGCTCAAAGCAAATCTTCAGCAG 602

    288 e.................GluTrp.................... 290
                           :::|||
    603 ACTCAGGAGAAATAGATGACAGTGGCAAATAGAGAGGTTAACAGTAAAGA 652

    291 ..........ArgAspProGly.......................... 294
                |||||   |||
    653 GAAGTGAAAATAGAGATGAAGGAGATTAGAGAGGTCTGAAATGAGAGAGA 702

    295 .....................ThrGlyArgThrGly.............. 299
                                |||||||||:::
    703 TACTCTGAATGAGAAATATCTAGAGGAAGAACCAGTCGATCCGAGTCAGA 752
```


300 .....GlnLeuThrTrpThrArgLeuProG 308
753 CGAGAAAAAGAGTTCAGAAAACATGAGGCTGACCAAGACC..... 792
308 lnGlyPheLysAsnSerProThrIlePheAspGluAlaLeuHis...Arg 323
793 .....GAGACCAGAGTCAGAATCAGCAGCTGTGCTGTGAGACCAAGA 834
324 AspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnTy 340
835 GAGATGAGAAAAATGAAATGTCGGCCACCAAAGTCCGGAAAATTAGAAAA 884
340 rValAspAspLeu.....LeuLeuAlaGlyAlaThrL 351
885 CTTAGATNGTACCTGGCAACAAAAAAAAAAGCTCTTGGCT.....C 925
351 ysGlnAspCysLeu...GluGlyThrLysAlaLeuLeuLeuGluLeuSer 366
926 AAGATCAATGTGTTTAAGAGATAACAAAAGGGGTGCTAATACAG..... 969
367 AspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGl 383
970 .....AAGCTGAGTCCTTAGAAGAGTCC 992
383 uValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrG 400
993 GGTGGCCTACCTG.....TTGAAGC 1012
400 luAlaArgLysLysThrValValGlnIleProAlaProThrThrAlaLys 416
1013 AGCTAGAAAAAGAGACTGTGTTTCATACTCCTCCACCGACCAGTGCAAAA 1062
417 GlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIlePr 433
1063 CAAGCTAAAAAGTTCCTGGGCACTGCGGGCTTTTGCAGATTGTGGATTCC 1112
433 oGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysG 450
1113 AGGTTTGCTTAAT.....TAAAGAGATAAAC 1138
450 lyGlyPheSerTrpAlaProGluHisGln.....LysAlaPheAspAla 464
1139 AGCCCTTCGTATAGAGAAATAAAAAACAACCTTGGATGGCCTTGGATGCC 1188
465 IleLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValTh 481
1189 ATTAAGACTGCCCTGATGTTATCCCCAGCTATGGGACTCCTAGATGTGAC 1238
481 rLysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyV 498
1239 T.....GAGAACAAAGGTATTGTCAAAGAAG 1264
498 alLeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyr.LeuSe 514
1265 TTCTTACTCAGAGATTGGGACCCTAGAAAAGACCTGTGGCATACTTGT.. 1312
514 rLysLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaI 531
1313 .AAAAAATTAGACCTGGTGGCTGTAGGATGGCCTGCTTGTCTGCAC.... 1357
531 leAlaAlaValAla.IleLeuValLysAspAlaAspLysLeuThrLeuGl 547
1358 .....ATAGTGGCTTCTGGTCAAGGACGCAGATAAATTGACTCTGAG 1399
547 yGlnAsnIle.....ThrValIle.AlaProHis..... 556
1400 ACAAAACTTGGCACATGTCCTAAAAAGTGTGGTTCAGCCCCCATGACCGA 1449
557 .....AlaLeuGluAsnIleValArg.GlnProProAspArgT 569
1450 TGGCTGACTAATGCTCTTGAAAACATTATCCAACCGTCCCCCTGACCGAT 1499
569 rpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGlu 585
1500 GG.....ACACATTATCAGAGCTTTTTGTTGACTGAA 1531
586 ArgValThrPheAlaProProAlaAlaLeuAsnProAlaThr.LeuLeuP 602
1532 CGAGTGACCTTCGCTCCCCCTGCTGTCCTCAAT.....CTCACTACTGC 1575
602 roGluGluThrAspGluProValThrHisAspCysHisGlnLeuLeuIle 618
1576 CTGAGACTTCGCCT.....ACTCATCATTGTGCTGACATTCTGGCA 1616
619 GluGluThrGlyValArgLysAspLeuThrAspIleProLeuThrGly.. 634
1617 GAAGGAACTCATACTCGAAATGATCTGAAAGATCAGATCAGCCTTGGCCT 1666
635 GluValLeuThrTrpPheThrAspGlySerSerTyrValValGluGlyLy 651
1667 GAGAGTTTGAGCTGGTACACGGATGGCAGTAGCCTGGAGGTTGAGGGTAA 1716
651 sArgMetAlaGly...Ala..AlaValValAspGlyThrArgThrIleTr 666
1717 GCGGAAGGCGGGGACAGCAGTGCAGTGGTGGACAGAAAGCAAGTGATCTA 1766
666 pAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetA 683
1767 GGCCAGCAGCCTCCCTAAAGGGACTTCAGCCCTGAAAGCTAAACTTGTGG 1816
683 laLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyr 699
1817 CTTTAATACAAGCT..... 1830
700 ThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAla.IleT 716
1831 .....CTATATAT 1838
716 yrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLys 732
1839 ACAGACAAAGAGAGCTGTTGACATCTGCA..... 1867
733 GluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAl 749
1867 1867
749 aIleIleHisCysProGlyHisGln.LysAlaLysAspLeuIleSerArg 765
1868 .....GAAAAAGACCTAAGATGCTGTGGCTAAA 1895
766 GlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAs 782
1896 AAAAATCAGATGGCAAATCTAACCGCCCAGGCATCCTAAAGAGCAATGAT 1945
782 nLeuLeuProIleIleGluThrProLysAlaProGluProArgArgGlnT 799
1946 CCTG..... 1949
799 yrThrLeu.GluAspTrpGlnGluIleLysLysIleAspGlnPheSerGl 815
1950 ..ACAGTCTAAAGACTATCAAGTTATAGACAAATTA..... 1983
815 uThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProH 832
1983 1983
832 isLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeu 848
1983 1983
849 GlyThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLe 865
1984 .....AGACTGGTAAAAGAAACCCTGTATGACATAGT 2015
865 uArgLeuProGlyValAlaAspSerValValLysHisCysValProCysG 882

```
2016 AGAAACTGAAGGGAACACGAATATAGTCCTCTCA..... 2049

 882 lnLeuValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArg 898

2050 .....TGA 2052

 899 GlySerHisProGlyAlaHisTrpGluValAspPheThrGluValLysPr 915
2053 GAAGACAGACCTGACATCTACTGAAAGATAGACTTCACTGAAGAAAAAAT 2102

 915 oAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerG 932
2103 GTGTATG......AATACCTTCTGG......TTTTTGTGAACGTTCTCAA 2140

 932 lyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAla 948
2141 GATGGATAAAAGCTTTTCCTTATAAAACTGAGACTACTCAGATAGTCATC 2190

 949 LysLysIleLeuGluGluIlePheProArgPheGlyIleProLysValIl 965
2191 AAGAAGATTATTAAAAAT...TTTCCAAGGTTCGGAGTGCCAAAAGCAAT 2237

 965 eGlySerAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuA 982
2238 AGTGTCAGATAATGGTCCTGCCTTTGTTGCCCAGGTAAGTCAGGGTGTGG 2287

 982 laLysIleLeuGlyIleAspTrpLys.LeuHisCysAlaTyrArgProGl 998
2288 CCAAGTATTTAGAGGTCAAATACAAGATTG.....TACAGACCTCA 2328

 998 nSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuT 1015
2329 GAACTCAGGACAGATAGAAATAATAGATAAAACTCT.AAACAGACCTTGA 2377

1015 hrLysLeuThrThrGluThrGlyIleAsnAsp.TrpMetAlaLeuLeuPr 1031
2378 CAAAATTAATCCTAGAGACTGGC...ACAGACTTACTTGGTGCCCCTTCC 2424

1031 oPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrP 1048
2425 CCTTGCCCTATTTAAAACTGAGAATACTCCCTCTTGATAGGGTTTTACTC 2474

1048 roTyrGluLeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPhe 1064
2475 TTTTTAAGATCCTTTATGGGGCTCCTATGCCCATCACTGTCTTAAATAAT 2524

1065 AlaHisSerAlaAspValLeuLeu.SerGlnProLeuPheSerArgLeu. 1080
2525 GTGTTTAAA...CCTATGTTGTTATAATAATGATCTGTATACTAAGTTAT 2571

1081 LysAlaLeuGluTrpValArgGlnArgAla.....\.TrpLysGln.L 1094
2572 AAGGCTTACAGGTGGTGCAGAAAGAAGTCTGGTCACAACTGGCTACAGTG 2621

1094 euArgGluAlaTyrSerGlyGlyAspLeuGlnValProHisArgPheGln 1110
2622 AACGAGCTGGGTACCCCAAGGACATCTTACCAGTTCCAGCCAGA..... 2665

1111 ValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluTh 1127
2666 ......GATCTGATCTGCGTACACCTGCGTCATGCTGAGACCCTCAAGCC 2709

1127 rArgTrpLysGlyProTyrLeuValLeuLeuThrThrPro 1140
2710 TCACTAGAAGGGTCCCTGCCTAGTTCTGTTTACTAATCCG 2749
```

OM of: US-09-171-553B-6 to: EST:* out_format : pfs

Date: Feb 24, 2002 8:31 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09171553/runat_22022002_122825_11137/app_query.fasta_1.3312
-DB=EST -QFMT=fastap -SUFFIX=feb22std.rst -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09171553_@CGN1_1_9008 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-171-553B-6
Query length: 656
Database: EST:*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 3515.700000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_est1:AW657531	+	894.00	1242.03	5.1e-60	549	! AW657531 110922 MARC 1BOV Bos t
gb_gss:AZ970955	-	881.50	1221.53	7.1e-59	754	! AZ970955 2M0244H16F Mouse 10kb
gb_gss:BH059027	+	865.50	1197.69	1.5e-57	882	! BH059027 RPCI-24-326G3.TJ RPCI-
gb_gss:BH122287	+	849.50	1176.62	2.2e-56	774	! BH122287 RPCI-24-288P13.TJ RPCI
gb_gss:AZ351242	-	820.00	1136.40	3.9e-54	701	! AZ351242 1M0089L09F Mouse 10kb
gb_est2:BI151181	+	787.00	1088.55	1.8e-51	845	! BI151181 602916935F1 NCI_CGAP_I
gb_gss:AZ850764	-	776.50	1076.55	8.4e-51	641	! AZ850764 2M0152C18R Mouse 10kb
gb_gss:AZ813042	-	763.50	1058.65	8.4e-50	625	! AZ813042 2M0080E11F Mouse 10kb
gb_est2:BI106669	+	763.00	1054.22	1.5e-49	921	! BI106669 602893272F1 NCI_CGAP_I
gb_gss:AZ356870	-	740.50	1026.26	5.3e-48	644	! AZ356870 1M0098F09F Mouse 10kb
gb_gss:AZ376517	-	730.50	1012.82	3.0e-47	610	! AZ376517 1M0130J20F Mouse 10kb
gb_gss:AZ430249	-	729.00	1011.04	3.8e-47	591	! AZ430249 1M0214G11R Mouse 10kb
gb_gss:AZ337993	+	727.50	1007.69	5.8e-47	673	! AZ337993 1M0068F24R Mouse 10kb
gb_gss:AZ642336	+	721.50	1000.05	1.5e-46	624	! AZ642336 1M0505M18F Mouse 10kb
gb_gss:AZ741412	+	719.50	996.73	2.4e-46	659	! AZ741412 RPCI-24-74C13.TJ RPCI-
gb_gss:AZ327473	-	717.00	993.11	3.7e-46	668	! AZ327473 1M0050K14R Mouse 10kb
gb_gss:AZ814037	+	713.50	988.76	6.5e-46	632	! AZ814037 2M0081N06R Mouse 10kb
gb_gss:AZ739547	-	697.00	965.08	1.4e-44	676	! AZ739547 RPCI-24-73L17.TJ RPCI-
gb_gss:AZ375836	-	696.50	966.11	1.2e-44	565	! AZ375836 1M0129J20F Mouse 10kb
gb_gss:AZ499035	+	693.00	961.77	2.1e-44	534	! AZ499035 1M0336P05R Mouse 10kb
gb_est2:BI109841	+	687.50	950.36	9.0e-44	787	! BI109841 602901065F1 NCI_CGAP_M
gb_gss:AZ417785	-	683.50	947.66	1.3e-43	583	! AZ417785 1M0193B08R Mouse 10kb
gb_est2:BF608155	+	676.00	933.76	7.6e-43	833	! BF608155 MY1_001169 Mouse 9-day
gb_est1:BE569478	+	673.00	928.89	1.4e-42	895	! BE569478 601328494F1 NCI_CGAP_M
gb_gss:AZ620796	-	671.50	929.87	1.2e-42	650	! AZ620796 1M0453I12R Mouse 10kb
gb_gss:AZ836232	+	670.00	928.67	1.5e-42	592	! AZ836232 2M0131E01F Mouse 10kb
gb_gss:AZ416306	-	668.50	926.37	2.0e-42	605	! AZ416306 1M0191D01R Mouse 10kb
gb_est2:BF159914	+	659.00	911.99	1.2e-41	680	! BF159914 601767582F1 NCI_CGAP_I
gb_gss:BH054319	-	659.00	911.39	1.3e-41	724	! BH054319 RPCI-24-398E11.TV RPCI
gb_gss:AZ336017	-	658.50	912.10	1.2e-41	625	! AZ336017 1M0066P05F Mouse 10kb
gb_est2:BF181106	+	649.00	894.63	1.1e-40	969	! BF181106 601805164F1 NCI_CGAP_M
gb_gss:AZ764389	+	648.50	898.40	7.1e-41	609	! AZ764389 1M0560M15F Mouse 10kb
gb_gss:AZ647666	+	642.00	888.28	2.6e-40	679	! AZ647666 1M0514J12F Mouse 10kb
gb_est2:BI106591	+	626.50	865.97	4.5e-39	728	! BI106591 602893174F1 NCI_CGAP_I
gb_est2:BI158931	+	625.50	864.74	5.3e-39	716	! BI158931 602921838F1 NIH_CGAP_M
gb_gss:AZ989131	+	624.50	864.18	5.7e-39	656	! AZ989131 2M0272E22F Mouse 10kb
gb_est2:BI082444	+	622.00	859.41	1.0e-38	750	! BI082444 602877669F1 NCI_CGAP_M
gb_gss:AZ349077	-	619.00	859.26	1.1e-38	493	! AZ349077 1M0085K21R Mouse 10kb
gb_gss:AZ960032	-	618.00	855.29	1.8e-38	644	! AZ960032 2M0227K23R Mouse 10kb
gb_gss:AZ397096	-	615.50	851.35	3.0e-38	675	! AZ397096 1M0161J22R Mouse 10kb
gb_gss:AZ996293	-	609.50	844.67	6.9e-38	566	! AZ996293 2M0282L19F Mouse
gb_est2:BF786434	+	606.00	836.44	2.0e-37	801	! BF786434 602112966F1 NCI_C
gb_gss:AZ887402	-	605.00	838.24	1.6e-37	575	! AZ887402 RPCI-24-176N10.TJ
gb_gss:BH055839	+	602.00	833.32	3.0e-37	620	! BH055839 RPCI-24-93P16.TVB
gb_gss:AZ018624	-	601.00	831.70	3.7e-37	635	! AZ018624 RPCI-23-382C22.TV

seq_name: gb_est1:AW657531

seq_documentation_block:
LOCUS AW657531 549 bp mRNA EST 25-APR-2001
DEFINITION 110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW657531
VERSION AW657531.1 GI:7423429
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 549)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR PRimers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 94 row: E column: 12
Seq primer: ATTTAGGTGACACTATAG.
FEATURES Location/Qualifiers
source 1. .549
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 145 a 144 c 121 g 139 t
ORIGIN

alignment_scores:
Quality: 894.00 Length: 183
Ratio: 5.109 Gaps: 1
Percent Similarity: 95.628 Percent Identity: 90.164

alignment_block:
US-09-171-553B-6 x AW657531 ..

Align seg 1/1 to: AW657531 from: 1 to: 549

```
    289 SerProGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGl 305
        |||||||||||||||||||||||||||||||||||||||||||||||||||
      3 TCCCCCGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCA 52

    305 nGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerC 322
        ||||||||||||||||||||||||||||||||||||||:::|||||||
```

```
 53 GGGAGCTTTCCAAGCCATCAACTCCACCGACCCCGATGCCAGTTCTTCTT 102

322 ysTrpLeuCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLys 338
    ||||||||||||||||||||||||||||||||||||||||||||||||:::
103 GTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGAATGGCTAGA 152

339 GluArgLysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGl 355
    |||   ||||||||||||||||||||||||||||||||||||||||||||
153 GAAGGGAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGG 202

355 ySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysI 372
    ||||||||||||||||||||||||||||||||||||||||||||||||||
203 GTCCCGAAATAAGCTCACCCTCACTGAAGTTTCCGGGAAGGGGACATGCA 252

372 leGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValVal 388
    ||||||||||||||||||||||||||||||||||   ||||||:::|||
253 TAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCAATAGTACTATGGTT 302

389 TyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTr 405
    ||||||||||||   ||||||||||||||||||||||||||||||   ||
303 TATGAGCAGGCCTTAGAAAATCAGTATTTAGTACCTGGTTATAACTGATG 352

405 pTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheA 422
    ||||||||||||||||||||||||||||||||||||||||:::|||||||
353 GTGGGCATGCAATACTGGGTTAACTCCCTGTGTTTCTACCACAGTCTTTA 402

422 snGlnSerLysAspPheCysValMetValGlnIleValProArgValTyr 438
    ||||||||||||||||||||||||||||||||:::|||||   |||:::
403 ACCAATCCAAAGATTTCTGTGTCATGGTCCAGCTCGTCCCCTGGGTACAC 452

439 TyrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnAr 455
    ||||||||||||||||||||:::|||||||||||||||||||   :::||
453 TACCATCCTGAGGAAGTAGTCATCGATGAGTATGACTATCGGCCTACCCG 502

455 gProLysArgGluProValSer.LeuThrLeuAlaValMetLeuGly 470
    |   ||||||||||||||:: ||||||||||||||:::||||||
503 CTCAAAAAGAGAGCCCGTGACCCTTAACCCTAGCCGTTATACTAGGA 549

seq_name: gb_gss:AZ970955

seq_documentation_block:
LOCUS       AZ970955      754 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0244H16F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0244H16 F, DNA sequence.
ACCESSION   AZ970955
VERSION     AZ970955.1  GI:13842182
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 754)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0244  row: H  column: 16
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 754.
FEATURES             Location/Qualifiers
     source          1. .754
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0244H16"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      226 a    160 c    150 g    218 t
ORIGIN


alignment_scores:
            Quality:  881.50              Length:    244
              Ratio:  4.025                 Gaps:      3
 Percent Similarity:  89.754   Percent Identity: 67.623

alignment_block:
 US-09-171-553B-6 x AZ970955/rev   ..

 Align seg 1/1  to reverse of: AZ970955  from: 1  to: 754

406 TrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAs 422
    ||||||||||||||||||||   ||||||||||||   |||   |||||
751 TGGGCTTGCAATACCGGTCTCATTCCTTGTGTGTCTATGTCTGATTTTAA 702

422 nGlnSerLysAspPheCysValMetValGlnIleValProArgValTyrT 439
    |:::|||||||||||||||:::::|||||:::::|||||:::   |
701 TAGTTCCAAAGATTTCTGCATTTTGGTTCAGCTTATTCCTAGACTCCTGT 652

439 yrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArg 455
    |||||   ::::::      ||||||:::::::::::|||      |||
651 ATCATGATGATAGCTCCTTTTTAGACAAATTTGAGCATCGGGTC...CGC 605

456 ProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGl 472
    :::|||||||||:::::::|||||||||:::|||:::||||||||||||
604 TGGAGAAGAGAGCCCATTACCTTAACTTTGACAGTTCTATTAGGATTAGG 555

472 yThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProG 489
    |:::|||:::|||||||||||||||||||||||||||||      ||||
554 AGTAGCGGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAAGACCCCCC 505

489 ln.GlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAs 505
    ||       |||            ||||||:::|||||||||      ||
504 AATACTAATGAA............GAACTACGTGCAGCTATGGATGTTGA 467

505 pLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrS 522
    |||||||:::::::::::::|||:::::::::|||||||||||||||||
466 TCTTAGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAACTT 417

522 erLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPhe 538
    ||||||||||||||||||||||||||||||||||||||||||||||||||
416 CCCTGTCCGAAGTGGTGCTACAGAATAGAAGGGGATTAGACTTATTATTC 367
```

```
     539 LeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTy 555
         |||:::||||||||||||||||||||||||||||||||||||||||||||
     366 CTTAAAGAAGGAGGACTCTGTGCTGCCCTAAAAGAAGAATGTTGTTTTTA 317

     555 rValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysL 572
         ||||||||||||||||:::|||:::||||||||   |||||||||:::: 
     316 TGTTGACCATTCAGGAGTAATCAAAGATTCTATGGCCAAACTTAGAGAAC 267

     572 ysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGlu 588
         ::|||:::   |||:::||||||||||||:::::||||||||||||||| 
     266 GCCTAGATACACGTAAAAGGGAAAGAGAAAGCCAACAAGGATGGTTCGAA 217

     589 GlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuTh 605
         :::|||||||||:::||||||||||:::|||||||||||||||:::::::: 
     216 AGCTGGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACCATAGC 167

     605 rGlyProLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuI 622
         :||||||||:::::|||:::||||||||||   |||||||||:::: 
     166 AGGACCTTTGATTACTCTTATGCTTTTGCTTACTTTTGGCCCCTGCATCC 117

     622 leAsnArgPheValAlaPheValArgGluArgValSerAlaValGlnIle 638
         ::|||:::::|||||||||:::|||||||||:::::|||||||||::: 
     116 TTAATAAGTTAGTAGCTTTTATTAGAGAAAGGATAAATGCAGTACAGGTT 67

     639 MetValLeuArgGlnGlnTyrGlnGlyLeu 648
         ||||||||||||||||||||:::   ||| 
      66 ATGGTACTAAGGCAACAATATCGGGTCCTT 37

seq_name: gb_gss:BH059027

seq_documentation_block:
LOCUS       BH059027      882 bp    DNA             GSS       18-JUL-2001
DEFINITION  RPCI-24-326G3.TJ RPCI-24 Mus musculus genomic clone RPCI-24-326G3,
            DNA sequence.
ACCESSION   BH059027
VERSION     BH059027.1  GI:14868646
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-326G3.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 326  row: G  column: 3
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .882
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-326G3"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      248 a    183 c    201 g    250 t
ORIGIN


alignment_scores:
           Quality:  865.50              Length:    309
             Ratio:  3.652                 Gaps:      6
 Percent Similarity:  76.699   Percent Identity: 58.252

alignment_block:
 US-09-171-553B-6 x BH059027   ..

 Align seg 1/1  to: BH059027  from: 1  to: 882

     297 GlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSe 313
         |||||||||||||||:::||||||:::|||||||||   |||:::|||||| 
       2 GGGCAAAGACTGTTCAACCTGATTAGAGGAGCCTTTTATGCCCTTAACAG 51

     313 rThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyP 330
         |:::|||||||||||||||:::   |||||||||   |||||||||||||| 
      52 CTCGGATCCAGATGCCACTACATACTGCTGGTTATTCCTATCATCAGGCC 101

     330 roProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLys 346
         ||||||||   ||||||:::|||   ::::::   ||||||   |||::: 
     102 CTCCCTACTCTGAAGGGGTTGCTTTTAATGAAGATTTTAATAAAACAAGT 151

     347 GluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeu.. 362
         :::|||          |||:::||||||:::   :::||||||||||||  
     152 AGTCATGCC...TTATGCTCATGGGGAACAGGAAAGAAATTAACCCTGTA 198

     363 ....ThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProProS 378
             :::      |||:::         |||   |||      ||||||: 
     199 CTGAAGTGTCAGGAGCAGCCCTTGTCTCTGTACAGGGACACCTCCTCCTG 248

     378 erHisGlnHisLeuCysTyrSerThr.ValValTyrGluGlnAlaSerGl 394
         ::|||||||||||||||      |||          :::::::::::  
     249 CCCACCAACACCTATGTGGTCGGACTTAATTCTGTATCCAGGACAAATAC 298

     394 uAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrG 411
         |||   ||||||||||||          ||||||||||||||||||||  
     299 CAACTATTACCTTGTGCCATATCCAGTTGGTTGGTGGGCCTGTAACACAG 348

     411 lyLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPhe 427
         |||||||||||||||||||||||||||||:::   |||   ||||||  
     349 GGCTCACCCCATGTGTATCAACTAGTGTCTTTGACCCCTCTGGTGATTTC 398

     428 CysValMetValGlnIleValProArgValTyrTyrHisProGluGluVa 444
         |||:::||||||:::::::::   |||:::||||||:::|||   :::  
     399 TGTATCATGGTTCATCTTCTATCTCGCATGTATTATTATCCTGCCTCCCG 448

     444 lValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProV 461
          :::   |||||||||      :::      :::   ||||||:::||||  
     449 CCTGGAAGATGAATATATCAGTAAG......AAGTTGAAAAGAGATCCAG 492

     461 alSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyVal 477
         |||||:::|||||||||:::::::::|||:::|||   :::   ||||||  
     493 TCTCCATTACTCTAGCTATGTTAATGGGGGTCGGTCTTACAGA.GGAGTA 541

     478 GlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGl 494
         |||   |||   |||||||||:::   |||   |||   ::::::  
     542 GGAATGGGTTTAGCAGCCTTGGTTGAGGGTAGACAAGGAATTCACACCTT 591

     494 yLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysG 511
          |||   |||         |||:::::|||||||      |||::::  
     592 CTTAAGAGAA.........GCTATAAATGAAGATTTAAGCATGCTACAAA 632

     511 luSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValVal 527
         ::|||||:::   ||||||:::|||||||||||||||||||||||||||  
     633 AATCTGTTGATGCATTAGAGAAATCCTTGACGTCCCTGTCAGAAGTAGTA 682
```

```
      528 LeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLe 544
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      683 TTACAAAACAGAAGAGGTTTAGATTTATTGTTTTTAAGGGAAGGTGGACT 732

      544 uCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyA 561
          ||||||||||||||||||||||||||||||||||:::||||||:::|||
      733 ATGTGCAGCCTTAAAAGAAGAATGCTGCTTCTATGCAGACCACACGGGAA 782

      561 laIleArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArg 577
            :::|||||||||||||:::|||||||||::::::|||||||||:::
      783 TAGTCAGAGATTCTATGCAGAAATTGAGGGAAAGACTGGAAAGGGAA... 829

      578 ArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSe 594
                                 |||||||::   |||::::::   ||
      830 .......................TGGTTCAAGGTGTGGTACTCCTAATC 855

      594 rProTrpMetThrThrLeuLeuSer 602
          |||||||||||||||||||||||||
      856 TCCATGGATGACCACTTTACTTTCT 880

seq_name: gb_gss:BH122287

seq_documentation_block:
LOCUS       BH122287      774 bp    DNA             GSS       19-JUL-2001
DEFINITION  RPCI-24-288P13.TJ RPCI-24 Mus musculus genomic clone RPCI-24-288P13
            , DNA sequence.
ACCESSION   BH122287
VERSION     BH122287.1  GI:14965799
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-288P13.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 288  row: P  column: 13
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-288P13"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      221 a    162 c    172 g    219 t
ORIGIN
```

```
alignment_scores:
           Quality:  849.50              Length:    262
             Ratio:  3.776                Gaps:      2
 Percent Similarity:  85.878   Percent Identity: 62.214

alignment_block:
 US-09-171-553B-6 x BH122287   ..

 Align seg 1/1  to: BH122287  from: 1  to: 774

     354 TrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyTh 370
         :::|||:::::::::|||||||||||:::   |||||||||:::|||
       2 TTTGGGGAGAAAACAAAGTTGACTCTGGCAGCAGTTTCAGGAAGAGGGCT 51

     370 rCysIleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrV 387
         |||:::|||::::::|||   :::      |||||||||   :::|||
      52 TTGTTTGGGCCAGGTACCTCAGGATAAAGGGCACCTCTGTAATCAGACCC 101

     387 alValTyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsn 403
               :::::::::::::::::::|||||||||||||||      :::
     102 AGAACATCCAGTCTAGCAAAAGTGGTCAGTATCTAGTGCCCCCCCTAGAC 151

     404 ArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerVa 420
             ||||||||||||||||||||||||||||||||||||   |||||
     152 ACAGTGTGGGCTTGCAATACCGGTCTCACTCCTTGTGTGTCTATGTCTGT 201

     420 lPheAsnGlnSerLysAspPheCysValMetValGlnIleValProArgV 437
         |||||||:::|||||||||||||||||:::::|||||:::::|||||:
     202 TTTTAATAGTTCCAAAGATTTCTGCATTTTGGTTCAGCTTATTCCTAGAC 251

     437 alTyrTyrHisProGluGluValValLeuAspGluTyrAspTyrArgTyr 453
         ::   ||||||   :::::::::   ||||||:::::::::::::|||
     252 TCCTGTATCATGATGATAGCGCCTTTTTAGACAAATTTGAGCATCGGGTC 301

     454 AsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGl 470
             |||   :::|||||||||||:::|||||||||||||||:::|||||
     302 ...CGCTGGAGAAGAGAGCCCGTTACCTTAACTTTGGCAGTTCTATTAGG 348

     470 yLeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrG 487
         |||||||:::|||:::||||||||||||||||||||||||||||||||||
     349 ATTAGGAGTAGCGGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAAGA 398

     487 lyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThr 503
           ||||||      |||            ||||||:::|||||||||
     399 CCCCCCAATACTATGAA.............GAACTACGTGCAGCTATGGAT 436

     504 GluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLe 520
         |||||||||:::::::::::|||::::::::|||||||||||||||
     437 GTTGATCTTAGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTT 486

     520 uThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuL 537
         |||||||||||||||||||||||||||||||||||||||||||||||||
     487 AACTTCCCTGTCCGAAGTGGTGCTACAAAATAGAAGGGGATTAGACTTAT 536

     537 euPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCys 553
         ||||||||:::|||   |||||||||||||||||||||||||||||||
     537 TATTCCTTAAAGAACGAGGACTCTGTGCTGCCCTAAAAGAAGAATGTTGA 586

     554 PheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuAr 570
         ||||||||||||||||||   :::|||:::|||||||||   ||||||||
     587 TTTTATGTTGACCATTCATGAGTAATCAAAGATTCTATGGCCAAACTTAG 636

     570 gLysLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpP 587
         |:::::|||:::   |||:::|||||   |||:::::|||||||||||
     637 AGAACGCCTAGATATACGTGAAAGAGAAATAGAAAGCCAACAAGGATGGT 686

     587 heGluGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAla 603
         |||||:::|||   |||:::|||||||||:::|||||||||||||||:::
     687 TCGAAAGCTGGGTTAATAAGTCCCCTTGGCTCACCACTCTTCTCTCCACC 736

     604 LeuThrGlyProLeuValValLeuLeuLeuLeuLeu 615
         ::::::|||||||||:::::::|||:::||||||||
```

```
    737 ATAGCAGGACCTTTGATTACTCTTATGCTTTTGCTT 772

seq_name: gb_gss:AZ351242

seq_documentation_block:
LOCUS       AZ351242      701 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0089L09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0089L09 F, DNA sequence.
ACCESSION   AZ351242
VERSION     AZ351242.1  GI:10430479
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0089  row: L  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 701.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0089L09"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      211 a    149 c    134 g    206 t      1 others
ORIGIN


alignment_scores:
            Quality:  820.00            Length:    234
              Ratio:  3.923               Gaps:      4
 Percent Similarity:  89.316   Percent Identity: 67.949

alignment_block:
 US-09-171-553B-6 x AZ351242/rev   ..

 Align seg 1/1  to reverse of: AZ351242  from: 1  to: 701

     416 ValSerThrSerValPheAsnGlnSerLysAspPheCysValMetValGl 432
         ||||||   |||||||||||||:::|||||||||||||||||:::::|||||
     700 GTGTCTATGTCTGTTTTTAATAGTTCCAAAGATTTCTGCATTTTGGTTCA 651

     432 nIleValProArgValTyrTyrHisProGluGluValValLeuAspGluT 449
         |:::::::|||||:::   ||||||   ::::::      ||||||   :
     650 ACTTATTCCTAGACTCCTGTATCATGATGATAGCTCCTTTTTAGACANAT 601

     449 yrAspTyrArgTyrAsnArgProLysArgGluProValSer.LeuThrLe 465
         :::::::::|||      |||   :::||||||||||||:: ||||||||
     600 TTGAGCATCGGGTC...CGCTGGAGAAGAGAACCCGTTACCCTTAACTTT 554

     465 uAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGlyThrA 482
         ||||||||:::|||||||||||||:::|||||||||||||||||||||||
     553 GGCAGTTCTATTAGGATTGGGAGTAGCGGTTGGAGTAGGTACAGGAACTG 504

     482 laAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeu 498
         ||||||||||||      ||||||      |||              ||||||
     503 CTGCCTTAATTAAGACCCCCCAATACTATGAA.....GAACTA 466

     499 HisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAs 515
         :::||||||||||      ||||||||||:::::::::::|||::::::::
     465 CGTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAACCAA 416

     515 nLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgA 532
         :||||||||||||||||||||||||||||||||||||||||||||||||||
     415 ATTAGAAGAATCTTTAACTTCCCTGTCCGAAGTGGTGCTACAGAATAGAA 366

     532 rgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeu 548
         |||||||||||||||||||||||:::   ||||||||||||||||||||
     365 GGGGATTAGACTTATTATTCCTTAAA...GGAGGACTCTGTGCTGCCCTA 319

     549 LysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSe 565
         ||||||||||||||||||||||||||||||||:::|||:::||||||
     318 AAAGAAGAGTGTTGTTTTTATGTTGACCATTCAGGAGTAATCAAAGATTC 269

     565 rMetAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArgGluA 582
         ||||   |||||||||:::::|||:::   |||:::||||||||||||:
     268 TATGGCCAAACTTAGAGAACGCCTAGATATACGTAAAAGAGAAAGAGAAA 219

     582 laAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMetThr 598
         :::::|||||||||||||||:::|||||||||:::|||||||||:::|||
     218 GCCAACAAGGATGGTTCGAAAGCTGGTTTAATAAGTCCCCTTGGCTCACC 169

     599 ThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLe 615
         ||||||||||||:::::::::|||||||||:::::|||:::|||||||||
     168 ACTCTCCTCTCCACCATAGCAGGACCTTTAATTACACTTATGCTTTTGCT 119

     615 uThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluA 632
         ||||   |||||||||:::::|||:::::::|||||||||:::|||||||
     118 TACTTTTGGCCCCTGCATCCTTAATAAGTTAGTAGCTTTTATTAGAGAAA 69

     632 rgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 648
         ||:::::|||||||||:::||||||||||||||||||||:::   |||
      68 GGATAAATGCAGTACAGGTTATGGTACTAAGGCAACAATATCGGGTCCTT 19

seq_name: gb_est2:BI151181

seq_documentation_block:
LOCUS       BI151181      845 bp    mRNA            EST       05-JUL-2001
DEFINITION  602916935F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5067273 5',
            mRNA sequence.
ACCESSION   BI151181
VERSION     BI151181.1  GI:14611182
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11181  row: f  column: 10
            High quality sequence stop: 774.
FEATURES              Location/Qualifiers
     source           1. .845
                      /organism="Mus musculus"
                      /strain="CZECH II (feral)"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5067273"
                      /clone_lib="NCI_CGAP_Lu29"
                      /tissue_type="spontaneous tumor, metastatic to mammary.
                      Stem cell origin."
                      /lab_host="DH10B"
                      /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigator
                      providing samples: Gilbert Smith, NIH"
BASE COUNT      225 a    188 c    216 g    215 t      1 others
ORIGIN


alignment_scores:
           Quality:  787.00              Length:    288
             Ratio:  3.422                 Gaps:      8
 Percent Similarity:  79.861   Percent Identity: 58.681

alignment_block:
 US-09-171-553B-6 x BI151181   ..

 Align seg 1/1  to: BI151181  from: 1  to: 845

     308 PheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLe 324
         |||   |||:::|||   |||||||||:::|||||:::::::|||||||
       2 TTTTATGCCCTTAACAGAACTGATCCAAGCGCTACTGAGGACTGCTGGCT 51

     324 uCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluArgL 341
         |||||||||   ||||||||||||||||||||:::|||   :::
      52 ATGCCTGTCCTTGGGTCCGCCTTATTATGAAGGAATCGCCTTCAATGGAG 101

     341 ysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySerArg 357
           ||||||   |||:::::|||   :::::::|||:::|||||:::
     102 ATTTCAACAGAACCAGCAGCCAT...ACTTCCTGCTCTTGGGGTACAGGA 148

     358 AsnLysLeuThrLeuThrGluValSerGlyLys......GlyThrCysIl 372
         :::|||||||||||||||||||||||:::::      |||   |||||
     149 CAAAAACTGACCCTGACTGAAGTATCCGCGAGGAATCCAGGTCTCTGTAT 198

     372 eGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValT 389
         ||||      |||   :::|||:::|||||||||   :::
     199 AGGTACCCCACCTTCCACTCACAAACACCTATGCGGACAAATTCAGTCCG 248

     389 yrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrp 405
         ::::::::::::   |||   ||||||||||::::          |||
     249 TGTCCAGAACGGAAGCTAATTACTATCTTGTACCTTCCCCGGTTGGATGG 298

     406 TrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAs 422
         |||||||||||||||||||||||||||||||||||||:::|||||||||
     299 TGGGCTTGCAATACAGGACTTACTCCCTGTGTATCAACTAAGGTTTTTAA 348

     422 nGlnSerLysAspPheCysValMetValGlnIleValProArgValTyrT 439
         |:::|||      ||||||||||||:::|||:::::::||||||||||||||
     349 TTCATCTCATGA.TTTTGTGTCATGATCCAGCTGTTACCCCGCGTATATT 397

     439 yrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArg 455
         ||||||||   :::   :::       |||   :::|||       :::|||
     398 ATCACCCTGCATCCAGTTTA......GAAGAAAGCTATGCTGGCCGGCGG 441

     456 ProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGl 472
         ||||||||||||:::::|||||||||||||:::::::::|||:::||
     442 TCAAAAAGAGAACCAATTACTTTAACCCTGGCTGCATTCATGGGAATAGG 491

     472 yThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProG 489
         |   ||||||||||||||||||||||:::::|||||||||   |||   |
     492 TATGGCAGTAGGAGTGGGGACGGGAGTGTCAGCTTTGATAGAAGGAAGAC 541

     489 lnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAsp 505
         ||            |||:::   :::|||:::   |||::::::||||||
     542 AG...........GGAATTCAGTCTTTGAGGGATGCTGTCAATGAAGAC 579

     506 LeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSe 522
         |||   |||:::::::::|||:::::   ||||||:::|||||||||||
     580 CTAGCGGCAATAGAGAAGTCCATTGACGCTTTAGAAAAATCTTTGACCTC 629

     522 rLeuSerGluValValLeuGlnAsnArgArgGly.LeuAspLeuLeuPhe 538
         |||||||||||||||||||||||||||||||||||| |||   |||||||||
     630 CCTGTCTGAGGTAGTTTTACAGAACAGGAGAGGTCTTGCATTTGTTGTTC 679

     539 LeuArgGluGlyGlyLeuCysAlaAlaLeuLysGlu.GluCysCysPheT 555
         |||:::|||||||||||||||||||||||||||||| ||||||||||||
     680 CTACAGGAAGGAGGACTGTGTGCTGCCCTTAAAGAAAGAGTGCTGCTTCT 729

     555 yrValAspHisSerGlyAlaIleArg.AspSerMetAsnLysLeuArgLy 571
         ||:::|||||||:::|||   :::||| |||||||||   |||||||||::
     730 ATGCAGATCATACAGGAATAGTTAGAAGACTCTATGCCGAAACTGAGGAG 779

     571 sLys...LeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpP 587
         :|||      ::::::   :::|||||||||:::|||      |||
     780 AAAGCATTCGAGCGAGGGACACCGGGAACGGGATGCTCACGGGGGTGGTG 829

     587 heGluGly 589
           :::|||
     830 AGTCCGGG 837

seq_name: gb_gss:AZ850764

seq_documentation_block:
LOCUS       AZ850764      641 bp    DNA             GSS       21-FEB-2001
DEFINITION  2M0152C18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0152C18 R, DNA sequence.
ACCESSION   AZ850764
VERSION     AZ850764.1  GI:13036088
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
```

```
          Plate: 0152  row: C  column: 18
          Seq primer: CACACAGGAAACAGCTATGACC
          Class: plasmid ends
          High quality sequence stop: 641.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0152C18"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      197 a    134 c    123 g    187 t
ORIGIN


alignment_scores:
            Quality:  776.50            Length:    218
              Ratio:  3.982               Gaps:      2
 Percent Similarity:  89.450   Percent Identity: 66.514

alignment_block:
 US-09-171-553B-6 x AZ850764/rev   ..

 Align seg 1/1  to reverse of: AZ850764  from: 1  to: 641

     416 ValSerThrSerValPheAsnGlnSerLysAspPheCysValMetValGl 432
         ||||||   ||||||||||||:::|||||||||||||||:::::::||||
     639 GTGTCTATGTCTGTTTTTAATAGTTCCAAAGATTTCTGCATTTTGGTTCA 590

     432 nIleValProArgValTyrTyrHisProGluGluValValLeuAspGluT 449
         |:::::::||||||:::   ||||||   ::::::      ||||||:::: 
     589 GCTTATTCCTAGACTCCTGTATCATGATGATAGCTCCTTTTTAGACAAAT 540

     449 yrAspTyrArgTyrAsnArgProLysArgGluProValSerLeuThrLeu 465
         :::::::|||      |||   :::||||||||||||:::|||||||||
     539 TTGAGCATCGGGTC...CGCTGGAGAAGAGAACCCGTTACCTTAACTTTG 493

     466 AlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGlyThrAl 482
         |||||||:::|||||||||||:::|||:::||||||||||||||||||||
     492 GCAGTTCTATTAGGATTGGGAGTAGCGGCTGGAGTAGGTACAGGAACCGC 443

     482 aAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuH 499
         ||||||||||      ||||||      |||             ||||||:
     442 TGCCTTAATTAAGACCCCCCAATACTATGAA.....GAACTAC 405

     499 isAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAsn 515
         ::|||||||||      |||||||||:::::::::::|||:::::::::
     404 GTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAACCAAA 355

     516 LeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgAr 532
         ||||||||||||||||||||||||||||||:::|||||||||||||||
     354 TTAGAAGAATCTTTAACTTCCCTGTCCGAAGTGTTGCTACAGAATAGAAG 305

     532 gGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuL 549
         |||||||||||||||||||||:::|||||||||||||||||||||||||
     304 GGGATTAGACTTATTATTCCTTAAAGAAGGAGGACTCTGTGCTGCCCTAA 255

     549 ysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSer 565
         ||||||||||||||||||||||||||||||||:::|||:::||||||
     254 AAGAGGAATGTTGTTTTTATGTTGACCATTCAGGAGTAATCAAAGATTCT 205

     566 MetAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArgGluAl 582
         |||   |||||||||:::::|||:::      :::||||||||||||::
     204 ATGGCCAAACTTAGAGAACGCCTAGATATACTTAAAAGAGAAAGAGAAAG 155

     582 aAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMetThrT 599
         ::::|||||||||||||||:::|||||||||:::   ||||||:::||||
     154 CCAACAAGGATGGTTCGAAAGCTGGTTTAATAAGTTCCCTTGGCTCACCA 105

     599 hrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeu 615
         ||||||||||:::::::::|||||||||:::::|||:::|||||||||
     104 CTCTCCTCTCCACCATAGCAGGACCTTTAATTACACTTATGCTTTTGCTT 55

     616 ThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluAr 632
         |||   |||||||||:::::|||:::::::||||||||||:::|||||||
      54 ACTTTTGGCCCCTGCATCCTTAATAAGTTGGTAGCTTTTATTAGAGAAAG 5

     632 gVal 633
         |:::
       4 GATA 1

seq_name: gb_gss:AZ813042

seq_documentation_block:
LOCUS       AZ813042      625 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0080E11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0080E11 F, DNA sequence.
ACCESSION   AZ813042
VERSION     AZ813042.1  GI:12982881
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0080  row: E  column: 11
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 625.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0080E11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
```

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 191 a 128 c 120 g 186 t
ORIGIN

alignment_scores:
Quality: 763.50 Length: 207
Ratio: 4.083 Gaps: 2
Percent Similarity: 90.338 Percent Identity: 69.565

alignment_block:
US-09-171-553B-6 x AZ813042/rev ..

Align seg 1/1 to reverse of: AZ813042 from: 1 to: 625

```
407 AlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGl 423
    |||||||||:::|||||||||||||||||||||   ||||||||||||::
607 GCTTGCAATAACGGTCTCACTCCTTGTGTGTCTATGTCTGTTTTTAATAG 558

423 nSerLysAspPheCysValMetValGlnIleValProArgValTyrTyrH 440
    :|||||||||||||||:::::|||||:::::|||||||:::   ||||
557 TTCTAAAGATTTCTGCATTTTGGTTCAGCTTATTCCCAGACTCTTGTATC 508

440 isProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgPro 456
    ||   ::::::      ||||||:::::::::::|||      |||
507 ATGATGATAGCTCTTTTTTAGATAAATTTGAACATCGGGTC...CGCTGG 461

457 LysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyTh 473
    |||||||||||||||:::|||||||||||||||::::|||||||||||::
460 AAAAGAGAACCCGTTACCTTAACTTTGGCAGTTCTATTAGGATTGGGAGT 411

473 rAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnG 490
    :|||:::|||||||||||||||||||||||||||||      ||||||
410 AGCAGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAAGACCCCCCAAT 361

490 lnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeu 506
        |||          |||||||:::|||||||||      ||||||
360 ACTATGAA...........GAACTACGTGCAGCTATGGATATTGATCTT 323

507 ArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLe 523
    |||:::::::::::::|||:::::::::|||||||||||||||||||||
322 AGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAACTTCCCT 273

523 uSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuA 540
    ||||||||||||||||||||||||||||||||||||||||||||||||:
272 GTCTGAAGTGGTGCTGCAAAATAGAAGGGGATTAGACTTATTATTCCTTA 223

540 rgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrVal 556
    ::|||||||||||||||||||||||||||||||||||||||||||||||
222 AAGAAGGAGGACTCTGTGCTGCCTTAAAAGAAGAATGTTGTTTTTATGTT 173

557 AspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLe 573
    ||||||||||||:::||:::||||||||   |||||||||::::::||
172 GACCATTCAGGAGTAATCAAAGATTCTATGGCTAAACTTAGAGAACGCTT 123

573 uGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyT 590
```

```
    |:::   |||:::|||||||||||:::::|||||||||||||||:::|
122 AGATATACGTAAAAGAGAAAGAGAGAGCCAACAAGGATGGTTCGAAAGCT 73

590 rpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGly 606
    ||||||||:::||||||||||:::|||||||||||||||:::::::::|||
 72 GGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACTATAGCAGGA 23

607 ProLeuValValLeuLeuLeu 613
    ||||||:::::::|||:::|||
 22 CCTTTAATTACACTTATGCTT 2
```

seq_name: gb_est2:BI106669

seq_documentation_block:

```
LOCUS       BI106669      921 bp    mRNA            EST       26-JUN-2001
DEFINITION  602893272F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5038336 5',
            mRNA sequence.
ACCESSION   BI106669
VERSION     BI106669.1  GI:14557562
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11105  row: p  column: 17
            High quality sequence stop: 805.
FEATURES             Location/Qualifiers
     source          1. .921
                     /organism="Mus musculus"
                     /strain="CZECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5038336"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      249 a    211 c    229 g    232 t
ORIGIN
```

alignment_scores:
Quality: 763.00 Length: 304
Ratio: 3.206 Gaps: 10
Percent Similarity: 78.289 Percent Identity: 54.276

alignment_block:
US-09-171-553B-6 x BI106669 ..

Align seg 1/1 to: BI106669 from: 1 to: 921

```
298 GlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSerTh 314
    ||||||:::|||:::|||::::::|||||||||   |||:::|||   ||
  4 CAGAGGATGTTTAACCTAGTGAGAGGAGCCTTTTATGCCCTTAACAGAAC 53

314 rAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProP 331
    |||||||:::||||||::::::|||||||||||||||||||||||||||||
```

```
 54 TGATCCAAGCGCTACTGAGGACTGCTGGCTATGCCTGTCCTCGGGTCCGC 103

331 roTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGlu 347
    |||||||||||||||:::|||   :::   :::||||||   |||:::::: 
104 CTTATTATGAAGGAATCGCCTTCAATGGAAATTTCAACAGAACCAGCAGC 153

348 HisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThrGl 364
    |||   ::::::|||:::|||||:::   :::|||||||||||||||| 
154 CAT...ACTTCCTGCTCTTGGGGTACAGGACAAAAACTGACCCTGACTGA 200

364 uValSerGlyLys......GlyThrCysIleGlyLysAlaProProSerH 379
    |||||||::::::      |||   |||||||||      |||   :::| 
201 AGTATCCGCGAGGAATCCAGGTCTCTGTATAGGTACCCCACCTTCCACTC 250

379 isGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsn 395
    ||:::|||||||||   :::            ::::::::::::   ::: 
251 ACAAACACCTATGCGGACAAATTCAGTCCGTGTCCAGAACGGAAGCTGAT 300

396 GlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLe 412
       ||||||||||::         ||||||||||||||||||||||||| 
301 TACTATCTTGTACCTTCCCCGGTTGGATGGTGGGCTTGCAATACAGGACT 350

412 uThrProCysValSerThrSerValPheAsnGlnSerLysAspPheCysV 429
    |||||||||||||||||||:::|||||||||:::|||   ||||||   | 
351 TACTCCCTGTGTATCAACTAAGGTTTTTAATTCATCTCATGATTTTCGTG 400

429 alMetValGlnIleValProArgValTyrTyrHisProGluGluValVal 445
    |||||:::|||:::::::|||||||||||||||||||||   :::   ::: 
401 TCATGATCCAGCTGTTACCCCGTGTATATTATCACCCTGCATCCAGTTTA 450

446 LeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSe 462
       |||:::|||      |||       |||   |||||||||||:::::: 
451 GAAGACAGCTATGCGGGCCGG......CGGTCAAAAAGAGAACCAATTAC 494

462 rLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyT 479
    :|||||||||||:::::::::|||:::|||   ||||||||||||||| 
495 TTTAACCCTGGCTGCATTCATGGGAATAGGTATGGCAGTAGGAGTGGGGA 544

479 hrGlyThrAlaAlaLeu.Ile.ThrGlyProGlnGlnLeuGluLysGlyL 495
    |||||:::::::|||::: ||| :::|||   |||            |||: 
545 CGGGAGTGTCAGCTTTCGATAGAACGGAAGACAG.............GGAA 582

495 euGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGlu 511
    ::   :::|||:::   |||::::::|||||||   |||:::::::::: 
583 TTCAGTCTTTGAGGGATGCTGTCAATGAAGACCTAGCGGCAATAGAGAAG 632

512 SerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLe 528
    |||::::::   ||||||:::|||||||||||||||||||||||||||| 
633 TCCATTGACGCTTTAGAAAAATCTTTGACCTCCCTGTCTGAGGTAGTTTT 682

528 uGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuC 545
    ||||||||||||||||||||            |||:::         | 
683 ACAGAACAGGAGAGGTCTTGATTCGTCGTCCTAAAGACAGGAGGACTGTT 732

545 ysAlaAlaLeuLysGluGluCysCysPheTyr.ValAspHisSerGly.A 561
    |||||||||||||||||||||||||||||||| :::|||||:::|||| 
733 GTGCTGCCCTTAAAGAAGAGTGCTGCTTCTATTGCAGATCATACAGGAAT 782

561 laIleArgAspSer.MetAsnLysLeuArgLysLysLeuGluArgArg.. 576
      :::|||||||| |||:::::|||   :::   ::::::   :::    
783 TAGTTAGAGACTCCTATGCAGAACCTGGACAGAACGATTTCAAGCGAAGG 832

577 .ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnA 593
       ||||||   :::|||:::::::|||         |||:::   :::| 
833 AACACGGGAACCGGATGCCCAACGGGGTAGCTGAGTCGGGTTTGAATCCA 882

593 rg 593
    || 
883 GA 884

seq_name: gb_gss:AZ356870

seq_documentation_block:
```

```
LOCUS       AZ356870      644 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0098F09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0098F09 F, DNA sequence.
ACCESSION   AZ356870
VERSION     AZ356870.1  GI:10470556
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 644)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0098  row: F  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 644.
FEATURES             Location/Qualifiers
     source          1. .644
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0098F09"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      185 a    136 c    133 g    190 t
ORIGIN


alignment_scores:
             Quality:  740.50              Length:    215
               Ratio:  3.897                 Gaps:      2
 Percent Similarity:  88.372   Percent Identity: 66.977

alignment_block:
 US-09-171-553B-6 x AZ356870/rev   ..

 Align seg 1/1  to reverse of: AZ356870  from: 1  to: 644
```

```
 390 GluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTr 406
     :::::::::::::::::::|||||||||||||||||      :::      ||
 639 CAGTCTAGTAAAAGTGGTCAGTATCTAGTGCCCCCCTCAGACACAGTATG 590

 406 pAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnG 423
     |||||||||||||||||||||||||||||||||||   ||||||||||||:
 589 GGCTTGCAATACCGGTCTCACTCCTTGTGTGTCTATGTCTGTTTTTAATA 540

 423 lnSerLysAspPheCysValMetValGlnIleValProArgValTyrTyr 439
     ::|||||:::|||||:::::::|||||:::::::|||||:::   |||
 539 GTTCCAAAAATTTCTGCATTTTGGTTCAGCTTATTCCTAGACTCCTGTAT 490

 440 HisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgPr 456
     |||   ::::::      ||||||:::::::::::::|||      |||
 489 CATGATGATAGCTCCTTTTTAGACAAATTTGAGCATCGGGTC...CGCTG 443

 456 oLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyT 473
      :::||||||||||||:::|||||||||||||||   ||||||||||||:
 442 GAGAAGAGAACCCGTTACCTTAACTTTGGCAGTTTA.TTAGGATTGGGAG 394

 473 hrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGln 489
     ::|||:::||||||||||||||||||||||||||||      ||||||
 393 TAGCGGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAAGACCCCCCAA 344

 490 GlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLe 506
           ||||||            |||:::|||||||||      |||||
 343 TACTATGAAAAA............CTACGTGCAGCTATGGATGTTGATCT 306

 506 uArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerL 523
     ||||:::::::::::::|||:::::::::|||||||||||||||||||||
 305 TAGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAACTTCCC 256

 523 euSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeu 539
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 255 TGTCCGAAGTGGTGCTACAGAATAGAAGGGGATTAGACTTATTATTCCTT 206

 540 ArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrVa 556
     :::||||||||||||||||||||||||||||||||||||||||||||||||
 205 AAAGAAGGAGGACTCTGTGCTGCCCTAAAAGAAGAATGTTGTTTTTATGT 156

 556 lAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysL 573
     |||||||||||||:::|||:::|||||||||   |||||||||:::::::|
 155 TGACCATTCAGGAGTAATCAAAGATTCTATGGCCAAACTTAGAGAACGCC 106

 573 euGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGly 589
     ||:::   |||:::|||||||||||||:::::::|||||||||||||:::
 105 TAGATATACGTAAAAGAGAAAGAGAAAGCCAACAAGGATGGTTCGAAAGC 56

 590 TrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeu 604
     |||||||||:::|||||||||:::|||||||||||||||||:::::
  55 TGGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACCATA 11

seq_name: gb_gss:AZ376517

seq_documentation_block:
LOCUS       AZ376517      610 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0130J20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0130J20 F, DNA sequence.
ACCESSION   AZ376517
VERSION     AZ376517.1  GI:10490217
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0130  row: J  column: 20
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 610.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0130J20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt-end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      177 a    131 c    120 g    182 t
ORIGIN


alignment_scores:
            Quality:  730.50              Length:    204
              Ratio:  3.970                 Gaps:      2
 Percent Similarity:  90.196   Percent Identity: 67.157

alignment_block:
 US-09-171-553B-6 x AZ376517/rev   ..

 Align seg 1/1  to reverse of: AZ376517  from: 1  to: 610

 431 ValGlnIleValProArgValTyrTyrHisProGluGluValValLeuAs 447
     ||||||:::::::||||||:::   ||||||   ::::::      |||||
 608 GTTCAGCTTATTCCTAGACTCCTGTATCATGATGATAGCTCCTTTTTAGA 559

 447 pGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSerLeuT 464
     |:::::::::::::|||      |||   :::|||||||||||:::||||
 558 CAGATTTGAGCATCGGGTC...CGCTGGAGAAGAGAGCCCGTTACCTTAA 512

 464 hrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGly 480
     ||||||||||:::||||||||||||:::|||:::|||||||||||||||||
 511 CTTTGGCAGTTCTATTAGGATTAGGAGTAGCGGCTGGAGTAGGTACAGGA 462

 481 ThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyGl 497
     ||||||||||||||    ||||||      |||            ||
 461 ACCGCTGCCTTAATTAAGACCCCCCAATACTATGAA.....GA 424

 497 uLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValS 514
     ||||:::|||||||||      |||||||||:::::::::::|||::::
```

```
     423 ACTACGTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAA 374

     514 erAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsn 530
         :::::|||||||||||||||||||||||||||||||||||||||||||||||
     373 CCAAATTAGAAGAATCTTTAACTTCCCTGTCCGAAGTGGTGCTACAGAAT 324

     531 ArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAl 547
         ||||||||||||||||||||||||||:::|||||||||||||||||||||
     323 AGAAGGGGATTAGACTTATTATTCCTTAAAGAAGGAGGACTCTGTGCTGC 274

     547 aLeuLysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArgA 564
         ||||||||||||||||||||||||||||||||||||||||:::|||:::|
     273 CCTAAAAGAAGAATGTTGTTTTTATGTTGACCATTCAGGAGTAATCAAAG 224

     564 spSerMetAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArg 580
         ||||||||   |||||||||:::::::|||:::   |||:::|||||||||
     223 ATTCTATGGCCAAACTTAGAGAACGCCTAGATATACGTAAAAGAGAAAGA 174

     581 GluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMe 597
         |||::::::|||||||||||||||:::|||||||||::|||||||||::
     173 GAAAGCCAACAAGGATGGTTCGAAAGCTGGTTTAATAAGTCCCCTTGGCT 124

     597 tThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuL 614
         :|||||||||||||||:::::::::|||||||||::::::|||:::||||
     123 CACCACTCTCCTCTCCACCATAGCAGGACCTTTGATTACACTTATGCTTT 74

     614 euLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArg 630
         ||||||||   |||||||||:::::::||::::::|||||||||:::|||
      73 TGCTTACTTTTGGCCCCTGCATCCTTAATAAGTTAGTAGCTTTTATTAGA 24

     631 GluArgValSer 634
         ||||||::::::
      23 GAAAGGATAAAT 12

seq_name: gb_gss:AZ430249

seq_documentation_block:
LOCUS       AZ430249      591 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0214G11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0214G11 R, DNA sequence.
ACCESSION   AZ430249
VERSION     AZ430249.1  GI:10554262
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0214  row: G  column: 11
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 591.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0214G11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      165 a    125 c    115 g    186 t
ORIGIN


alignment_scores:
            Quality:  729.00            Length:    194
              Ratio:  4.096               Gaps:      1
 Percent Similarity:  91.753   Percent Identity: 71.134

alignment_block:
 US-09-171-553B-6 x AZ430249/rev   ..

 Align seg 1/1  to reverse of: AZ430249  from: 1  to: 591

     455 ArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLe 471
         |||   |||||||||||||||:::|||||||||||||||||:::||||||||
     588 CGCTGGAAAAGAGAACCCGTTACCTTAACTTTGGCAGTTTTATTAGGATT 539

     471 uGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyP 488
         ||||:::|||:::|||||||||||||||||||||||||||||:::   |
     538 GGGAGTAGCAGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAACACCC 489

     488 roGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGlu 504
         |||||      |||           ||||||:::|||||||||
     488 CCCAATACTATGAA.....GAGCTACGTGCAGCTATGGATGTT 451

     505 AspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuTh 521
         ||||||||:::::::::::::|||:::::::::|||||||||||||||||
     450 GATCTTAGAACCATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAAC 401

     521 rSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuP 538
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     400 TTCCCTATCCGAAGTGGTGCTGCAAAATAGAAGGGGGTTAGACTTATTAT 351

     538 heLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPhe 554
         |||||:::||||||||||||||||   ||||||||||||||||||||
     350 TCCTTAAAGAAGGAGGACTCTGTGCTCCCTTAAAAGAAGAATGTTGTTTT 301

     555 TyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLy 571
         ||||||||||||||||||:::|||:::|||||||||   ||||||:::::
     300 TATGTTGACCATTCAGGAGTAATCAAAGATTCTATGGCTAAACTTAAAGA 251

     571 sLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheG 588
         :::::|||:::   |||:::|||||||||||||:::   ||||||||||||
     250 ACGCCTAGATATACGTAAAAGAGAAAGAGAAAGCCCACAAGGATGGTTTG 201

     588 luGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeu 604
         ||:::|||||||||:::||||||||||:::|||||||||||||||:::::
     200 AAAGCTGGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACTATA 151
```

```
     605 ThrGlyProLeuValValLeuLeuLeuLeuThrValGlyProCysLe 621
         :::||||||||||:::::::|||:::|||||||||||   |||||||||::
     150 GCAGGACCTTTAATTACACTTATGCTTTTGCTTACTTTTGGCCCATGCAT 101

     621 uIleAsnArgPheValAlaPheValArgGluArgValSerAlaValGlnI 638
         ::::|||:::::::|||||||||:::|||||||||:::::::|||||||||:
     100 CCTTAATAAGTTAGTAGCTTTTATTAGAGAAAGGATAAACGCAGTCCAAG 51

     638 leMetValLeuArgGlnGlnTyrGlnGlyLeu 648
         ::|||||||||||||||||||||:::   |||
      50 TTATGGTACTAAGGCAACAATATCGGGTCCTT 19

seq_name: gb_gss:AZ337993

seq_documentation_block:
LOCUS       AZ337993      673 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0068F24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0068F24 R, DNA sequence.
ACCESSION   AZ337993
VERSION     AZ337993.1  GI:10410826
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0068  row: F  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 673.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0068F24"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      197 a    120 c    167 g    188 t      1 others
ORIGIN


alignment_scores:
            Quality:  727.50              Length:    228
              Ratio:  3.638                Gaps:       3
 Percent Similarity:  87.719   Percent Identity: 59.211

alignment_block:
 US-09-171-553B-6 x AZ337993   ..

 Align seg 1/1  to: AZ337993  from: 1  to: 673

     420 ValPheAsnGlnSerLysAspPheCysValMetValGlnIleValProAr 436
         |||||||||:::|||   |||||||||||||||:::|||:::::|||||
           1 GTTTTTAATTCATCTCATGATTTTTGTGTCATGATCCAGCTGTTACCCCG 50

     436 gValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyrArgT 453
         |||||||||||||||   :::   :::      |||   :::|||
          51 CGTATATTATCACCCTGCATCCAGTTTA......GAAGAAAGCTATGCTG 94

     453 yrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeu 469
           :::|||   ||||||||||||:::::|||||||||||:::::::::
          95 GCCGGCGGTCAAAAAGAGAACCAATTACTTTAACCCTGGCTGCATTCATG 144

     470 GlyLeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleTh 486
         |||:::|||   ||||||||||||||||||:::::|||||||||
     145 GGAATAGGTATGGCAGTAGGAGTGGGGACGGGAGTGTCAGCTTTGATAGA 194

     486 rGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetT 503
          |||   |||            |||:::   :::|||:::   |||::::
     195 AGGAAGACAG.............GGAATTCAGTCTTTGAGGGATGCTGTCA 232

     503 hrGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSer 519
         ::|||||||||   |||:::::::::|||:::::   ||||||:::|||
     233 ATGAAGACCTAGCAGCAATAGAGAAGTCCATTGACGCTTTAGAAAAATCT 282

     520 LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLe 536
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     283 TTGACCTCCCTGTCTGAGGTAGTTTTACAGAACAGGAGAGGTCTTGATTT 332

     536 uLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysC 553
         ||||||||||:::||||||||||||||||||||||||||||||||||||||
     333 GTTGTTCCTAAAGGAAGGAGGACTGTGTGCTGCCCTTAAAGAAGAGTGCT 382

     553 ysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeu 569
         |||||||:::||||||:::|||   :::|||||||||||:::|||||||
     383 GCTTCTATGCAGATCATACAGGAATAGTTAGAGACTCTATGCAGAAACTG 432

     570 ArgLysLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTr 586
         |||::::::||||||:::|||:::||||||||||:::|||:::::|||||
     433 AGAGAAAGATTAGAGCAAAGGAAACGGGAACGGGATGCTCAACGGGGGTG 482

     586 pPheGluGlyTrpPheAsnArgSerPro...TrpMetThrThrLeuLeuS 602
         |||||||:::|||||:::   :::|||   |||:::|||:::|||:::|
     483 GTTTGAGTCGTGGTTTGAATCACAACCATCTTGGATAACTTCTTTAATTT 532

     602 erAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThrValGly 618
         |||||:::::|||||:::::::::::   |||      :::   |||
     533 CCGCTGTAGCCGGACCAATCCTTATGATATGCTTAGCTNTAGTTTTCGGC 582

     619 ProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAl 635
         ||||||:::|||||||||   :::||||||:::::::::::::::::::
     583 CCTTGTATAATAAATAGAGGAATGGCTTTCATCCAGAGTAAAATTGATAC 632

     635 aValGlnIleMetValLeuArgGlnGlnTyrGln 646
         :|||:::::||||||||||:::::|||||||||
     633 AGTAAAACTCATGGTTCTTCAAAGGCAATATCAA 666

seq_name: gb_gss:AZ642336
```

```
seq_documentation_block:
LOCUS       AZ642336      624 bp    DNA             GSS       14-DEC-2000
DEFINITION  1M0505M18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0505M18 F, DNA sequence.
ACCESSION   AZ642336
VERSION     AZ642336.1  GI:11768843
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0505  row: M  column: 18
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 624.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0505M18"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      203 a    116 c    131 g    174 t
ORIGIN


alignment_scores:
            Quality:  721.50            Length:    197
              Ratio:  3.964               Gaps:      2
 Percent Similarity:  92.386   Percent Identity: 70.051

alignment_block:
 US-09-171-553B-6 x AZ642336   ..

 Align seg 1/1  to: AZ642336  from: 1  to: 624
```

```
      460 ProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGl 476
          ||||||:::|||||||||||||||||:::|||||||||||||:::|||:::||
        3 CCCGTTACCTTAACTTTGGCAGTTCTATTAGGATTGGGAGTAGCAGCTGG 52

      476 yValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluL 493
          ||||||||||||||||||||||||||||      ||||||      |||
       53 AGTAGGTACAGGAACCGCTGCCTTAATTAAGACCCCCCAATACTATGAA. 101

      493 ysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeu 509
                     ||||||:::|||||||||      |||||||||:::::
      102 ..........GAACTACGTGCAGCTATGGATATTGATCTTAGAACTATA 140

      510 LysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluVa 526
          ::::::|||::::::::::||||||||||||||||||||||||||||||||
      141 GAACAGTCTATAACCAAATTAGAAGAATCTTTAACTTCCCTGTCCGAAGT 190

      526 lValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyG 543
          ||||||||||||||||||||||||||||||||||||||||:::|||||||
      191 GGTGCTGCAAAATAGAAGGGGATTAGACTTATTATTCCTTAAAGAAGGAG 240

      543 lyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHisSer 559
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      241 GACTCTGTGCTGCCTTAAAAGAAGAATGTTGTTTTTATGTTGACCATTCA 290

      560 GlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGluArgAr 576
          |||:::|||:::||||||||   ||||||||:::::|||:::   ||
      291 GGAGTAATCAAAGATTCTATGGCTAAACTTAGAGAACGCCTAGATATACG 340

      576 gArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnA 593
          |:::||||||||||||:::::||||||||||||||:::|||||||||:
      341 TAAAAGAGAAAGAGAAAGCCAACAAGGATGGTTTGAAAGCTGGTTTAATA 390

      593 rgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuVal 609
          ::||||||||:::|||||||||||||||:::::::::||||||||||:::
      391 AGTCCCCTTGGCTCACCACTCTCCTCTCCACTATAGCAGGACCTTTAATT 440

      610 ValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheVa 626
          :::|||:::||||||||||||   ||||||||:::::|||:::::||
      441 ACACTTATGCTTTGCTTACTTTGGCCCATGCATCCTTAATAAGTTAGT 490

      626 lAlaPheValArgGluArgValSerAlaValGlnIleMetValLeuArgG 643
          |||||||:::|||||||||:::::|||||||||:::|||||||||||||
      491 AGCTTTTATTAGAGAAAGGATAAACGCAGTCCAAGTTATGGTCCTAAGAC 540

      643 lnGlnTyrGlnGlyLeuLeuSerGlnGlyGluThrAspLeu 656
          ||||||||:::   :::|||:::::   |||:::::|||
      541 AACAATATCGG...GTCCTTCAGGAGGTTGAAAACTCGCTC 578

seq_name: gb_gss:AZ741412

seq_documentation_block:
LOCUS       AZ741412      659 bp    DNA             GSS       25-JAN-2001
DEFINITION  RPCI-24-74C13.TJ RPCI-24 Mus musculus genomic clone RPCI-24-74C13,
            DNA sequence.
ACCESSION   AZ741412
VERSION     AZ741412.1  GI:12518156
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-74C13.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
```

```
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 74  row: C  column: 13
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .659
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="RPCI-24-74C13"
                      /clone_lib="RPCI-24"
                      /sex="Male"
                      /cell_type="Spleen/Brain"
                      /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                      RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                      library was  cloned in the pTARBAC1 cloning vector at the
                      BamH1 sites using MboI  partially digested male C57BL/6J
                      DNA."
BASE COUNT      186 a    132 c    152 g    189 t
ORIGIN


alignment_scores:
            Quality:  719.50             Length:    220
              Ratio:  3.868                Gaps:      2
 Percent Similarity:  84.545   Percent Identity: 63.182

alignment_block:
 US-09-171-553B-6 x AZ741412   ..

 Align seg 1/1  to: AZ741412  from: 1  to: 659

     355 GlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCy 371
         |||   ::::::||||||||||||::   |||||||||::|||   ||
      10 GGGAGAAAACAGAAGTTGACTCTAGCAGCAGTTTCAGGAAGAGGGCTTTG 59

     371 sIleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValV 388
         |:::|||:::::|||            |||||||||   :::|||
      60 TCTGGGCCGGGTACCTCAGGTTAAAGGGCACCTCTGTAATCAGACCCAGA 109

     388 alTyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArg 404
               ::::::::::::::::::|||||||||||||||      :::
     110 ACATCCAGTCTAGCAAAAGTGGTCAGTATCTAGTGCCCCCTCTAGACACA 159

     405 TrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPh 421
             |||||||||||||||||||||||||||||||||   ||||||||
     160 GTGTGGGCTTGCAATACCGGTCTCACTCCTTGTGTGTCTATGTCTGTTTT 209

     421 eAsnGlnSerLysAspPheCysValMetValGlnIleValProArgValT 438
         ||||:::||||||||||||||||:::::|||||:::::|||||:::
     210 TAATAGTTCCAAAGATTTCTGCATTTTGGTTCAGCTTATTCCTAGACTCC 259

     438 yrTyrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsn 454
           ||||||   ::::::      ||||||:::::::::::|||
     260 TGTATCATGATGATAGCTCCTTTTTAGACAAATTTGAGCATCGGGTC... 306

     455 ArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLe 471
         |||   :::||||||||||||:::|||||||||||||||:::||||||||
     307 CGCTGGAGAAGAGAACCCGTTACCTTAACTTTGGCAGTTCTATTAGGATT 356

     471 uGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyP 488
         ||||:::|||:::|||||||||||||||||:::|||||||||      |
     357 GGGAGTAGCGGCTGGAGTAGGTACAGGAACCACTGCCTTAATTAAGACCC 406

     488 roGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGlu 504
         |||||      |||            ||||||:::|||||||||
```

```
     407 CCCAATACTATGAA...........GAACTACGTGCAGCTATGGATGTT 444

     505 AspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuTh 521
         ||||||||:::::::::::::|||::::::::::|||||||||||||||||
     445 GATCTTAGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAAC 494

     521 rSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuP 538
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     495 TTCCCTGTCCGAAGTGGTGCTACAGAATAGAAGGGGATTAGACTTATTAT 544

     538 heLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPhe 554
         |||||:::||||||||||||||||||||||||||||||||||||||||||
     545 TCCTTAAAGAAGGAGGACTCTGTGCTGCCCTAAAAGAAGAATGTTGTTTT 594

     555 TyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLy 571
         ||||||||||||||||||::|||:::|||||||||   |||||||||::
     595 TATGTTGACCATTCAGGAGTAATCAAAGATTCTATGGCCAAACTTAGAGA 644

     571 sLysLeuGlu 574
         ::::|||:::
     645 ACGCCTAGAT 654
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:37:03 ; Search time 6550.43 Seconds
(without alignments)
4849.227 Million cell updates/sec

Title: US-09-171-553B-9
Perfect score: 2956
Sequence: 1 tgctttttagggttaggaac.....aaaaaaaaaaaaaaaaaaaa 2956

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	472	16.0	549	10	AW657531	AW657531 110922 MA
c	2	383.6	13.0	564	11	BI182930	BI182930 UNL-P-FN-
c	3	358.6	12.1	602	11	BF712162	BF712162 MI-P-E6-a
c	4	293.6	9.9	754	13	AZ970955	AZ970955 2M0244H16
c	5	271.2	9.2	494	11	BI182742	BI182742 UNL-P-FN-
c	6	258.6	8.7	701	13	AZ351242	AZ351242 1M0089L09
	7	255	8.6	774	13	BH122287	BH122287 RPCI-24-2
c	8	254.4	8.6	591	13	AZ430249	AZ430249 1M0214G11
	9	253	8.6	632	13	AZ814037	AZ814037 2M0081N06
	10	246.8	8.3	624	13	AZ642336	AZ642336 1M0505M18
c	11	245.2	8.3	641	13	AZ850764	AZ850764 2M0152C18
c	12	242.6	8.2	493	13	AZ349077	AZ349077 1M0085K21
c	13	240.8	8.1	650	13	AZ620796	AZ620796 1M0453I12
c	14	240.6	8.1	625	13	AZ813042	AZ813042 2M0080E11
	15	234.6	7.9	592	13	AZ836232	AZ836232 2M0131E01
	16	232	7.8	534	13	AZ499035	AZ499035 1M0336P05
c	17	230.6	7.8	610	13	AZ376517	AZ376517 1M0130J20
c	18	229.4	7.8	644	13	AZ356870	AZ356870 1M0098F09
	19	228.6	7.7	609	13	AZ439189	AZ439189 1M0229N10
	20	228	7.7	730	11	BI185535	BI185535 UNL-P-FN-
c	21	227.6	7.7	565	13	AZ375836	AZ375836 1M0129J20
c	22	225	7.6	676	13	AZ739547	AZ739547 RPCI-24-7
	23	224	7.6	600	13	AZ987182	AZ987182 2M0269J09
	24	222.2	7.5	895	10	BE569478	BE569478 601328494
	25	221.6	7.5	597	13	AZ406484	AZ406484 1M0175C17
	26	221.6	7.5	882	13	BH059027	BH059027 RPCI-24-3
	27	220	7.4	576	13	AZ730616	AZ730616 RPCI-24-1
	28	219	7.4	834	13	BH124666	BH124666 RPCI-24-2
c	29	215.4	7.3	583	13	AZ417785	AZ417785 1M0193B08
	30	214.2	7.2	669	13	BH040768	BH040768 RPCI-24-2
c	31	214	7.2	554	11	BF703805	BF703805 MI-P-O2-a
	32	211.2	7.1	659	13	AZ741412	AZ741412 RPCI-24-7
c	33	211	7.1	485	11	BF703671	BF703671 MI-P-E4-a
	34	210.6	7.1	704	13	BH051865	BH051865 RPCI-24-2
	35	209.6	7.1	461	10	AL120896	AL120896 DKFZp762N
	36	209	7.1	640	13	BH058168	BH058168 RPCI-24-3
	37	209	7.1	679	13	AZ647666	AZ647666 1M0514J12
	38	208.8	7.1	636	13	AZ727116	AZ727116 RPCI-24-1
c	39	208	7.0	628	13	AZ406003	AZ406003 1M0175N06
	40	207.2	7.0	906	11	BI328061	BI328061 602979695
c	41	207	7.0	668	13	AZ327473	AZ327473 1M0050K14
c	42	200.6	6.8	631	11	BF702741	BF702741 MI-P-E3-a
	43	198.8	6.7	442	13	AZ597961	AZ597961 1M0412K22
c	44	198.2	6.7	625	13	AZ336017	AZ336017 1M0066P05
c	45	197.8	6.7	605	13	AZ416306	AZ416306 1M0191D01

ALIGNMENTS

RESULT 1
AW657531
LOCUS AW657531 549 bp mRNA EST 25-APR-2001
DEFINITION 110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW657531
VERSION AW657531.1 GI:7423429
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 549)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
PCR PRimers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 94 row: E column: 12

```
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT      145 a    144 c    121 g    139 t
ORIGIN


  Query Match             16.0%;  Score 472;  DB 10;  Length 549;
  Best Local Similarity   92.5%;  Pred. No. 1.7e-91;
  Matches  507;  Conservative    0;  Mismatches   40;  Indels    1;  Gaps    1;

Qy   1230 ctccccaggtgttcctgttaagacaggacagagactcttcagtctcatccagggagcttt 1289
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 CTCCCCCGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCAGGGAGCTTT 61

Qy   1290 ccaagccatcaactccaccgaccctgatgccacttcttcttgttggctttgtctatcctc 1349
          |||||||||||||||||||||||| ||||||| ||||||||||||||||||||||||||
Db     62 CCAAGCCATCAACTCCACCGACCCCGATGCCAGTTCTTCTTGTTGGCTTTGTCTATCCTC 121

Qy   1350 agggcctccttattatgaggggatggctaaagaaagaaaattcaatgtgaccaaagagca 1409
          |||||||||||||||||||| ||||||| |||| | ||||||||||||||||||||||||
Db    122 AGGGCCTCCTTATTATGAGGGAATGGCTAGAGAAGGGAAATTCAATGTGACCAAAGAGCA 181

Qy   1410 tagaaatcaatgtacatgggggtcccgaaataagcttaccctcactgaagtttccgggaa 1469
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    182 TAGAAATCAATGTACATGGGGGTCCCGAAATAAGCTCACCCTCACTGAAGTTTCCGGGAA 241

Qy   1470 ggggacatgcataggaaaagctcccccatcccaccaacacctttgctatagtactgtggt 1529
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||
Db    242 GGGGACATGCATAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCAATAGTACTATGGT 301

Qy   1530 ttatgagcaggcctcagaaaatcagtatttagtacctggttataacaggtggtgggcatg 1589
          |||||||||||||| ||||||||||||||||||||||||||||| | ||||||||||||
Db    302 TTATGAGCAGGCCTTAGAAAATCAGTATTTAGTACCTGGTTATAACTGATGGTGGGCATG 361

Qy   1590 caatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaagatttgtg 1649
          |||||||||||||| |||||||||| ||| ||||||| |||||||||||||||||| ||
Db    362 CAATACTGGGTTAACTCCCTGTGTTTCTACCACAGTCTTTAACCAATCCAAAGATTTCTG 421

Qy   1650 tgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtccttgatga 1709
          |||||||||||  |||||||| | ||  |||||||||||||||||||| ||| | |||||
Db    422 TGTCATGGTCCAGCTCGTCCCCTGGGTACACTACCATCCTGAGGAAGTAGTCATCGATGA 481

Qy   1710 atatgactatcggtataaccgaccaaaaagagaacccgtatccctt-accctagctgtaa 1768
           ||||||||||||  || |||  |||||||||| |||||  ||||| |||||||| || |
Db    482 GTATGACTATCGGCCTACCCGCTCAAAAAGAGAGCCCGTGACCCTTAACCCTAGCCGTTA 541

Qy   1769 tgctcgga 1776
          | || |||.
Db    542 TACTAGGA 549


RESULT   2
BI182930/c
LOCUS       BI182930      564 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-br-h-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-br-h-08-0-UNL 3', mRNA sequence.
ACCESSION   BI182930
VERSION     BI182930.1  GI:14657339
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized porcine ovarian follicles library
            Seq primer: M13 -29
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-br-h-08-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected during 7 days of the follicular phase
                     of the pig estrous cycle. This library was derived from
                     the library UNL-P-F2. The tag is a string of 5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track. The library was constructed as described
                     by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
                     , 1996.
                     TAG_LIB=UNL-P-FN
                     TAG_TISSUE=porcine ovarian follicles
                     TAG_SEQ=CACACT"
BASE COUNT      141 a    129 c    143 g    151 t
ORIGIN


  Query Match             13.0%;  Score 383.6;  DB 11;  Length 564;
  Best Local Similarity   84.6%;  Pred. No. 1.8e-72;
  Matches  477;  Conservative    0;  Mismatches    9;  Indels   78;  Gaps    1;

Qy   2467 cctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagggcctgcta 2526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    564 CCTGCTGGCTGCCAGTAAATAGGTAGAAGGTCACACTTCCTATTGTTCCAGGGCCTGCTA 505

Qy   2527 tcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctggattctgtaa 2586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    504 TCCTGGCCTAAGTAAGATAACAGGAAATGAGTTGACTAATCGCTTATCTGGATTCTGTAA 445

Qy   2587 aactgactggcaccatagaagaattgattacacattgacagccctagtgacctatctcaa 2646
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    444 AACTGACTGGCACCATAGAAGAACTGATTACACATTGACAGCCCTAGTGACCTATCTCAA 385

Qy   2647 ctgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagctattttaa 2706
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    384 CTGCAATCTGTCACTCTGCCCAGGAGCCCACGCAGATGCGGACCTCCAGAGCTATTTTAA 325

Qy   2707 aatgattggtccacggagcgcgggctctcgatattttaaaatgattggtccatggagcgc 2766
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    324 AATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAATGATTGGTCCACGGAGCGC 265

Qy   2767 g---------------------------------------------------------- 2767
          |
```

```
Db     264 GGGCTCTCGATATTTTAAAATGATTGGTCCACGGAGCGCGGGCTCTCGATATTTTAAAAT 205

Qy    2768 ------------------ggctctcgatattttaaaatgattggtttgtgacgcacagg 2808
                              ||||||||||||||||||||||||||||||||||||||||
Db     204 GATTGGTCCACGAAGCGCCGGCTCTCGATATTTTAAAATGATTGGTTTGTGACGCACAGG 145

Qy    2809 ctttgttgtgaaccccataaaagctgtcccgattccgcactcggggccgcagtcctctac 2868
           |||||||||||||||||||||||||||||| ||||||| ||  ||||||||||||||||
Db     144 CTTTGTTGTGAACCCCATAAAAGCTGTCCCGACTCCGCACACGCCGCCGCAGTCCTCTAC 85

Qy    2869 ccctgcgtggtgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtt 2928
           ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db      84 CCCTGCGTGGCCTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTT 25

Qy    2929 tgcatcaaaaaaaaaaaaaaaaaa 2952
           ||||||||||||||||||||||||
Db      24 TGCATCAAAAAAAAAAAAAAAAAA 1


RESULT   3
BF712162/c
LOCUS       BF712162      602 bp    mRNA            EST       02-JAN-2001
DEFINITION  MI-P-E6-acx-e-11-1-UM.s1 MI-P-E6 Sus scrofa cDNA clone
            MI-P-E6-acx-e-11-1-UM 3', mRNA sequence.
ACCESSION   BF712162
VERSION     BF712162.1  GI:12011595
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized fetus at gestational day 20 library cDNA Library
            Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
            Research Center, Department of Animal Science, University of
            Missouri-Columbia, 65211 Clone distribution: clones will be
            available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-E6-acx-e-11-1-UM"
                     /clone_lib="MI-P-E6"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-E6
                     library is derived from fetus at gestational day 20.  For
                     a detailed description of the  library from which this
                     clone was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/.
                     TAG_LIB=MI-P-E6
                     TAG_TISSUE=fetus at gestational day 20
                     TAG_SEQ=AGGAA"

BASE COUNT      141 a    146 c    152 g    162 t      1 others
ORIGIN


  Query Match           12.1%;  Score 358.6;  DB 11;  Length 602;
  Best Local Similarity   72.9%;  Pred. No. 4.4e-67;
  Matches  557;  Conservative    0;  Mismatches   45;  Indels  162;  Gaps    1;

Qy    2186 ccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttg 2245
           ||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     602 CCCTAGTGGTCCTGCTCCTGTTGCTTACAGTTGGGCCTTGCTTAATTAATAGGTTTGTTG 543

Qy    2246 cctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaag 2305
           ||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     542 CCTNTGTTAGAGAATGAGTGAGTGCAGTCCAGATCATGGTACTTAGGCAACAGTACCAAG 483

Qy    2306 gccttctgagccaaggagaaactgacctctagccttcccagttctaagattagaactatt 2365
           |||||| |||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     482 GCCTTCCGAGCCAAGGAGAAACTGACCTTTAGCCTTCCCAGTTCTAAGATTAGAACTATT 423

Qy    2366 aacaagacaagaagtggggaatgaaaggatgaaaatgcaacctaaccctcccagaaccca 2425
           |||||||||||||||||||||||||||||||||||| |||||||||| ||| ||||
Db     422 CACAAGACAAGAAGTGGGGAATGAAAGGATGAAAATGCAGCCTAACCCTCCAAGACCCCA 363

Qy    2426 ggaagttaataaaaagctctaaatgcccccgaattacagaccctgctggctgccagtaaa 2485
           |||| || | || ||| |||| | |  |
Db     362 AGAAGGTAGTGAAGAGCCCTAACTCCAGC------------------------------ 332

Qy    2486 taggtagaaggtcacacttcctattgttccagggcctgctatcctggcctaagtaagata 2545

Db     333 ------------------------------------------------------------ 332

Qy    2546 acaggaaatgagttgactaatcgcttatctggattctgtaaaactgactggcaccataga 2605

Db     333 ------------------------------------------------------------ 332

Qy    2606 agaattgattacacattgacagccctagtgacctatctcaactgcaatctgtcactctgc 2665
                       | |  | ||||||||||||||||||||||||||||||||||||| ||
Db     333 -----------TTCCTAAAAAGCCCTAGTGACCTATCTCAACTGCAATCTGTCACTCCGC 285

Qy    2666 ccaggagcccacgcagatgcggacctccggagctattttaaaatgattggtccacggagc 2725
           |||||||||| |||| |||||||||||||| || |   |||| |||||||||||||||
Db     284 CCAGGAGCCCGCGCAAATGCGGACCTCCGGAACTGTAGAAAAACGATTGGTCCACGGAGT 225

Qy    2726 gcgggctctcgatattttaaaatgattggtccatggagcgcgggctctcgatattttaaa 2785
           |||||||||||||| |||||||||||||||||| ||||| |||||||||||| |||||||
Db     224 GCGGGCTCTCGATGTTTTAAAATGATTGGTCCACGGAGCACGGGCTCTCGATGTTTTAAA 165

Qy    2786 atgattggtttgtgacgcacaggctttgttgtgaaccccataaaagctgtcccgattccg 2845
           |||||||||||| || || | ||||||||||||||||||||||||||||||||||||||||
Db     164 ATGATTGGTTTGCGAAGCGCGGGCTTTGTTGTGAACCCCATAAAAGCTGTCCCGATTCCG 105

Qy    2846 cactcggggccgcagtcctctacccctgcgtggtgtacgactgtgggccccagcgcgctt 2905
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     104 CACTCGGGGCCGCAGTCCTCTACCCCTGCGTGGCGTACGACTGTGGGCCCCAGCGCGCTT 45

Qy    2906 ggaataaaaatcctcttgctgtttgcatcaaaaaaaaaaaaaaa 2949
           |||||||||||||||||||||||||||||||||||||||||||||
Db      44 GGAATAAAAATCCTCTTGCTGTTTGCATCAAAAAAAAAAAAAAA 1


RESULT   4
AZ970955/c
LOCUS       AZ970955      754 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0244H16F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0244H16 F, DNA sequence.
ACCESSION   AZ970955
VERSION     AZ970955.1  GI:13842182
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 754)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0244  row: H  column: 16
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 754.
FEATURES             Location/Qualifiers
     source          1. .754
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0244H16"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      226 a    160 c    150 g    218 t
ORIGIN
```

Query Match 9.9%; Score 293.6; DB 13; Length 754;
Best Local Similarity 64.8%; Pred. No. 4.2e-53;
Matches 474; Conservative 0; Mismatches 244; Indels 14; Gaps 2;

```
Qy   1582 tgggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaaccaatccaaa 1641
          ||||| ||||||||  ||  | |  || ||||| || |  ||  |  || ||    ||||||
Db    751 TGGGCTTGCAATACCGGTCTCATTCCTTGTGTGTCTATGTCTGATTTTAATAGTTCCAAA 692

Qy   1642 gatttgtgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgaggaagtggtc 1701
          |||||  ||  |  |||| ||  |  | ||  ||  |     || |||  |||       |
Db    691 GATTTCTGCATTTTGGTTCAGCTTATTCCTAGACTCCTGTATCATGATGATAGCTCCTTT 632

Qy   1702 cttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatcccttacccta 1761
           | ||  ||| |||  |||||     |||    | |||||| ||| |  || | ||  |
Db    631 TTAGACAAATTTGAGCATCGG---GTCCGCTGGAGAAGAGAGCCCATTACCTTAACTTTG 575

Qy   1762 gctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctgccctgatc 1821
           | ||  |  | ||||||||    || | ||| ||||| ||||| || ||||||  | ||
Db    574 ACAGTTCTATTAGGATTAGGAGTAGCGGCTGGAGTAGGTACAGGAACCGCTGCCTTAATT 515

Qy   1822 acaggaccacagcagctagagaaaggacttggtgagctacatgcggccatgacagaagat 1881
          |     || ||  |   |           ||  || |||| ||| || |||    |  |||
Db    514 AAGACCCCCCAATACTAA----------TGAAGAACTACGTGCAGCTATGGATGTTGAT 466

Qy   1882 ctccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaa 1941
          ||  || |  || |  ||||| | | |||  |||||||  ||  | |||||  |||| |||
Db    465 CTTAGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAACTTCCCTGTCCGAA 406

Qy   1942 gtggttctacagaaccggaggggattagatctgctgtttctaagagaaggtgggttatgt 2001
          |||||  ||||||||  | ||||||||||  |  | || || | ||||||  ||  | |||
Db    405 GTGGTGCTACAGAATAGAAGGGGATTAGACTTATTATTCCTTAAAGAAGGAGGACTCTGT 346

Qy   2002 gcagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatcagagactcc 2061
          || ||| ||||||||||||||||| || ||||| || || |||||||  |||| ||| ||
Db    345 GCTGCCCTAAAAGAAGAATGTTGTTTTTATGTTGACCATTCAGGAGTAATCAAAGATTCT 286

Qy   2062 atgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccagggg 2121
          |||  ||| ||||||  ||    |||| |  |||  ||||||||||||      | || ||
Db    285 ATGGCCAAACTTAGAGAACGCCTAGATACACGTAAAAGGGAAAGAGAAAGCCAACAAGGA 226

Qy   2122 tggtttgaaggatggttcaacaggtctccttggatgaccaccctgctttctgctctgacg 2181
          |||||  ||| | |||||  || | |||  ||||||  |  |||||  ||  ||  ||  |  |  |
Db    225 TGGTTCGAAAGCTGGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACCATAGCA 166

Qy   2182 gggcccctagtagtcctgctcctgttacttacagttgggccttgcttaattaataggttt 2241
          || ||  |  |    ||  | || ||  |||||  ||||  ||  |||  |  ||||||  |||
Db    165 GGACCTTTGATTACTCTTATGCTTTTGCTTACTTTTGGCCCCTGCATCCTTAATAAGTTA 106

Qy   2242 gttgcctttgttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtac 2301
          || || ||| |||||||| |  | | ||||||  ||| | ||||||||  ||||||||  ||
Db    105 GTAGCTTTTATTAGAGAAAGGATAAATGCAGTACAGGTTATGGTACTAAGGCAACAATAT 46

Qy   2302 caaggccttctg 2313
          |  | |||||  |
Db     45 CGGGTCCTTCAG 34
```

```
RESULT   5
BI182742/c
LOCUS       BI182742      494 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-bo-e-05-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-bo-e-05-0-UNL 3', mRNA sequence.
ACCESSION   BI182742
VERSION     BI182742.1  GI:14657151
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized porcine ovarian follicles library
            Seq primer: M13 -29
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
```

```
                    Lines"
                    /db_xref="taxon:9823"
                    /clone="UNL-P-FN-bo-e-05-0-UNL"
                    /clone_lib="UNL-P-FN"
                    /dev_stage="ADULT"
                    /lab_host="DH10B (Life Technologies)"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                    library is a normalized library representing porcine
                    ovarian follicles, ranging between 2.0 to 10.0 mm in
                    diameter, collected during 7 days of the follicular phase
                    of the pig estrous cycle. This library was derived from
                    the library UNL-P-F2. The tag is a string of 5-6
                    nucleotides present between the Not I site and the
                    oligo-dT track. The library was constructed as described
                    by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
                    , 1996.
                    TAG_LIB=UNL-P-FN
                    TAG_TISSUE=porcine ovarian follicles
                    TAG_SEQ=CACACT"
BASE COUNT      118 a    117 c    125 g    134 t
ORIGIN


  Query Match             9.2%;  Score 271.2;  DB 11;  Length 494;
  Best Local Similarity   79.1%;  Pred. No. 3.1e-48;
  Matches  391;  Conservative    0;  Mismatches   58;  Indels   45;  Gaps    4;

Qy   2459 ttacagaccctgctggctgccagtaaataggtagaaggtcacacttcctattgttccagg 2518
           || |||  |||||| |||   | || |||||  |||| |    |       |  |  |
Db    449 TTCCAGGGCCTGCTCTCTGTATGAAACTAGGTGAAAGGAC---GTAAAATAGGCCCTTGA 393

Qy   2519 gcctgctatcctggcctaagtaagataacaggaaatgagttgactaatcgcttatctgga 2578
           |   | ||||||||||||||||||| |||||||||||| |||||||||||||||||||
Db    392 GTGCGTGATCCTGGCCTAAGTAAGATCACAGGAAATGAGTGGACTAATCGCTTATCTGGA 333

Qy   2579 ttctgtaaaactgactggcaccatagaagaattgattacacattgacagccctagtgacc 2638
           |||||||||||| || || |||||||||| |||||||||||| ||||||||||||
Db    332 TTCTGTAAAACTG-CTTGCGCCATAGAAGAAATGATTACACATTGAAAGCCCTAGTGACC 274

Qy   2639 tatctcaactgcaatctgtcactctgcccaggagcccacgcagatgcggacctccggagc 2698
           |||||||||||||||||||||||| |||||||||||| |||| ||||||||||||| | |
Db    273 TATCTCAACTGCAATCTGTCACTCCGCCCAGGAGCCCGCGCAAATGCGGACCTCCGCAAC 214

Qy   2699 tattttaaaatgattggtccacggagcgcgggctctcgatattttaaaatgattggtcca 2758
           | |   ||||||||||||||||
Db    213 TGTAGAAAAATGATTGGTCCAC------------------------------------ 190

Qy   2759 tggagcgcgggctctcgatattttaaaatgattggtttgtgacgcacaggctttgttgtg 2818
            |||| || |||||||||| |||||||||||||||||| || || | ||||||||||||
Db    191 -GGAGTGCCGGCTCTCGATGTTTTAAAATGATTGGTTTGCGAAGCGCGGGCTTTGTTGTG 133

Qy   2819 aaccccataaaagctgtcccgattccgcactcggggccgcagtcctctacccctgcgtgg 2878
           |||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||
Db    132 AACCCCATAAAAGCTGTCCCAATTCCGCACTTGGGGCCGCAGTCCTCTACCCCTGCGTGG 73

Qy   2879 tgtacgactgtgggccccagcgcgcttggaataaaaatcctcttgctgtttgcatcaaaa 2938
             ||||||  |||||||||||||||||||||||||||||||||||||||||||||||| |
Db     72 CCTACGAC--TGGGCCCCAGCGCGCTTGGAATAAAAATCCTCTTGCTGTTTGCATCAAGA 15

Qy   2939 aaaaaaaaaaaaaa 2952
           ||||||||||||||
Db     14 AAAAAAAAAAAAAA 1


RESULT   6
AZ351242/c
LOCUS       AZ351242      701 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0089L09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0089L09 F, DNA sequence.
ACCESSION   AZ351242
VERSION     AZ351242.1  GI:10430479
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0089  row: L  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 701.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0089L09"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      211 a    149 c    134 g    206 t      1 others
ORIGIN


  Query Match             8.7%;  Score 258.6;  DB 13;  Length 701;
  Best Local Similarity   64.6%;  Pred. No. 1.5e-45;
  Matches  455;  Conservative    0;  Mismatches  230;  Indels   19;  Gaps    4;

Qy   1611 tgtttccacctcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtccc 1670
           ||| || |  || || || ||    |||||||||| ||  |  |||| ||| |  | ||
Db    701 TGTGTCTATGTCTGTTTTTAATAGTTCCAAAGATTTCTGCATTTTGGTTCAACTTATTCC 642

Qy   1671 ccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcggtataaccg 1730
             || |    || ||| |||         |  | ||   || |||  |||||     |||
Db    641 TAGACTCCTGTATCATGATGATAGCTCCTTTTTAGACANATTTGAGCATCGGGT---CCG 585

Qy   1731 accaaaaagagaacccgtatcccttacccta-gctgtaatgctcggattagggacggccg 1789
             | ||||||||||||  |||||| | |  || ||  |  | ||||| ||    || |
Db    584 CTGGAGAAGAGAACCCGTTACCCTTAACTTTGGCAGTTCTATTAGGATTGGGAGTAGCGG 525
```

```
Qy   1790 ttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaaaggac 1849
          |||| ||||| ||||| || |||||| | || |     || |                ||
Db    524 TTGGAGTAGGTACAGGAACTGCTGCCTTAATTAAGACCCCCC-----------AATACT 477

Qy   1850 ttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgttagca 1909
          ||  || |||| ||| || |||   |  |||||  || |  || |  ||||| | | ||
Db    476 ATGAAGAACTACGTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAACCA 417

Qy   1910 acctagaagagtccctgacttctttgtctgaagtggttctacagaaccggaggggattag 1969
          |  ||||||| ||  | |||||  |||| |||||||| ||||||||  | ||||||||||
Db    416 AATTAGAAGAATCTTTAACTTCCCTGTCCGAAGTGGTGCTACAGAATAGAAGGGGATTAG 357

Qy   1970 atctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttct 2029
          |  |  | || || |   |||| ||  | ||||| ||| ||||||||||| ||||| || |
Db    356 ACTTATTATTCCTTA---AAGGAGGACTCTGTGCTGCCCTAAAAGAAGAGTGTTGTTTTT 300

Qy   2030 atgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagttagaga 2089
          |||| || || ||||||| |||| ||| || |||  ||| ||||||  ||    |||| |
Db    299 ATGTTGACCATTCAGGAGTAATCAAAGATTCTATGGCCAAACTTAGAGAACGCCTAGATA 240

Qy   2090 ggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctc 2149
            ||| ||| ||||||||     | || || ||||| ||| | ||||| || | ||| |
Db    239 TACGTAAAAGAGAAAGAGAAAGCCAACAAGGATGGTTCGAAAGCTGGTTTAATAAGTCCC 180

Qy   2150 cttggatgaccaccctgctttctgctctgacggggccctagtagtcctgctcctgttac 2209
          ||||| | ||||| || || ||  |  |  | || ||  || |   ||  | || || |
Db    179 CTTGGCTCACCACTCTCCTCTCCACCATAGCAGGACCTTTAATTACACTTATGCTTTTGC 120

Qy   2210 ttacagttgggccttgcttaattaataggtttgttgccttttgttagagaacgagtgagtg 2269
          ||||  |||| || ||| |  |||||| ||| || || ||| ||||||||  |  | | ||
Db    119 TTACTTTTGGCCCCTGCATCCTTAATAAGTTAGTAGCTTTTATTAGAGAAAGGATAAATG 60

Qy   2270 cagtccagatcatggtacttaggcaacagtaccaaggccttctg 2313
          |||| ||| | |||||||| |||||||| || |  | ||||| |
Db     59 CAGTACAGGTTATGGTACTAAGGCAACAATATCGGGTCCTTCAG 16


RESULT   7
BH122287
LOCUS       BH122287      774 bp    DNA             GSS       19-JUL-2001
DEFINITION  RPCI-24-288P13.TJ RPCI-24 Mus musculus genomic clone RPCI-24-288P13
            , DNA sequence.
ACCESSION   BH122287
VERSION     BH122287.1  GI:14965799
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-288P13.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 288  row: P  column: 13
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-288P13"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      221 a    162 c    172 g    219 t
ORIGIN


  Query Match             8.6%;  Score 255;  DB 13;  Length 774;
  Best Local Similarity   60.1%;  Pred. No. 8.5e-45;
  Matches  467;  Conservative    0;  Mismatches  295;  Indels   15;  Gaps    2;

Qy   1437 aaataagcttaccctcactgaagtttccgggaaggggacatgcataggaaaagctccccc 1496
          ||  ||| | || ||  | | |||||| || |  |||   ||  | ||  | |  || |
Db     13 AACAAAGTTGACTCTGGCAGCAGTTTCAGGAAGAGGGCTTTGTTTGGGCCAGGTACCTCA 72

Qy   1497 atcccaccaacacctttgctatagtactgtggtttatgagcaggcctcagaaaatcagta 1556
                |    ||||| ||  ||   ||   |        ||      |  |     ||||||
Db     73 GGATAAAGGGCACCTCTGTAATCAGACCCAGAACATCCAGTCTAGCAAAAGTGGTCAGTA 132

Qy   1557 tttagtacctggttataacaggtggtgggcatgcaatactgggttaacccccctgtgtttc 1616
          | |||| ||         |||    |||||| |||||||| ||  | || || ||||| ||
Db    133 TCTAGTGCCCCCCCTAGACACAGTGTGGGCTTGCAATACCGGTCTCACTCCTTGTGTGTC 192

Qy   1617 cacctcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtcccccgagt 1676
            |  || || || ||    ||||||||||| ||  |  |||| ||  |  | ||  || |
Db    193 TATGTCTGTTTTTAATAGTTCCAAAGATTTCTGCATTTTGGTTCAGCTTATTCCTAGACT 252

Qy   1677 gtactaccatcctgaggaagtggtccttgatgaatatgactatcggtataaccgaccaaa 1736
             || |||  |||     |   |  | ||  ||| |||  |||||     |||    |
Db    253 CCTGTATCATGATGATAGCGCCTTTTTAGACAAATTTGAGCATCGG---GTCCGCTGGAG 309

Qy   1737 aagagaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgt 1796
          |||||| |||||  || | ||  | || ||  |  | ||||||||    || | ||| ||
Db    310 AAGAGAGCCCGTTACCTTAACTTTGGCAGTTCTATTAGGATTAGGAGTAGCGGCTGGAGT 369

Qy   1797 aggaacagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtga 1856
          ||| ||||| || |||||| | || |     || ||  |    ||                ||
Db    370 AGGTACAGGAACCGCTGCCTTAATTAAGACCCCCCAATACTATGA-----------AGA 417

Qy   1857 gctacatgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctaga 1916
           |||| ||| || |||   |  |||||  || |  || |  ||||| | | |||  ||||
Db    418 ACTACGTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAACCAAATTAGA 477

Qy   1917 agagtccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgct 1976
          ||| ||  | |||||  |||| ||||||||| ||||| ||  | ||||||||||  |  |
Db    478 AGAATCTTTAACTTCCCTGTCCGAAGTGGTGCTACAAAATAGAAGGGGATTAGACTTATT 537

Qy   1977 gtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtaga 2036
          || || | |||| | ||  | ||||| ||| ||||||||||||||||| || |||||  ||
Db    538 ATTCCTTAAAGAACGAGGACTCTGTGCTGCCCTAAAAGAAGAATGTTGATTTTATGTTGA 597

Qy   2037 tcactcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcg 2096
           || ||| |||  |||| ||| || |||  ||| |||||| ||    |||| |  |||
Db    598 CCATTCATGAGTAATCAAAGATTCTATGGCCAAACTTAGAGAACGCCTAGATATACGTGA 657

Qy   2097 aagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggat 2156
          |||  |||| |||     | || || ||||| ||| | ||| | || | ||| ||||||  |
Db    658 AAGAGAAATAGAAAGCCAACAAGGATGGTTCGAAAGCTGGGTTAATAAGTCCCCTTGGCT 717

Qy   2157 gaccaccctgctttctgctctgacggggcccctagtagtcctgctcctgttacttac 2213
          ||||| || || ||  |  |  | || ||  |  |    ||  | || || |||||
Db    718 CACCACTCTTCTCTCCACCATAGCAGGACCTTTGATTACTCTTATGCTTTTGCTTAC 774
```

```
RESULT   8
AZ430249/c
LOCUS       AZ430249      591 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0214G11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0214G11 R, DNA sequence.
ACCESSION   AZ430249
VERSION     AZ430249.1  GI:10554262
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0214  row: G  column: 11
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 591.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0214G11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      165 a    125 c    115 g    186 t
ORIGIN


  Query Match             8.6%;  Score 254.4;  DB 13;  Length 591;
  Best Local Similarity   66.2%;  Pred. No. 1.2e-44;
  Matches  388;  Conservative    0;  Mismatches  186;  Indels   12;  Gaps    1;

Qy   1728 ccgaccaaaaagagaacccgtatccttaccctagctgtaatgctcggattagggacggc 1787
           |||    |||||||||||||  || | ||  | || ||  |  | |||||  ||    ||
Db    589 CCGCTGGAAAAGAGAACCCGTTACCTTAACTTTGGCAGTTTTATTAGGATTGGGAGTAGC 530

Qy   1788 cgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaaagg 1847
           | ||| ||||| ||||| || |||||| | || |     || ||  |    ||
Db    529 AGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAACACCCCCCAATACTATGAA----- 473

Qy   1848 acttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgttag 1907
                |||||| ||| || |||   |  |||||  || || || |  ||||| | |
Db    474 -------GAGCTACGTGCAGCTATGGATGTTGATCTTAGAACCATAGAACAGTCTATAAC 422

Qy   1908 caacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggaggggatt 1967
           |||  ||||||| ||  | |||||  | || ||||||| || || ||  | ||||| ||
Db    421 CAAATTAGAAGAATCTTTAACTTCCCTATCCGAAGTGGTGCTGCAAAATAGAAGGGGGTT 362

Qy   1968 agatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgctt 2027
           |||  |  | || || | |||||| ||  | |||||  |||||||||||||||||| ||
Db    361 AGACTTATTATTCCTTAAAGAAGGAGGACTCTGTGCTCCCTTAAAAGAAGAATGTTGTTT 302

Qy   2028 ctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagttaga 2087
           |||||  || || |||||||  |||| ||| || |||   || |||| | ||    ||||
Db    301 TTATGTTGACCATTCAGGAGTAATCAAAGATTCTATGGCTAAACTTAAAGAACGCCTAGA 242

Qy   2088 gaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtc 2147
           |  |||  ||| |||||||||       || || |||||||||| | ||||| || | |||
Db    241 TATACGTAAAAGAGAAAGAGAAAGCCCACAAGGATGGTTTGAAAGCTGGTTTAATAAGTC 182

Qy   2148 tccttggatgaccaccctgctttctgctctgacgggggcccctagtagtcctgctcctgtt 2207
           ||||||  | ||||| || || ||  || |  | || ||  || |    ||  | || ||
Db    181 CCCTTGGCTCACCACTCTCCTCTCCACTATAGCAGGACCTTTAATTACACTTATGCTTTT 122

Qy   2208 acttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacgagtgag 2267
           |||||  |||| || ||| |  |||||| ||| || || ||| |||||||| |  | |
Db    121 GCTTACTTTTGGCCCATGCATCCTTAATAAGTTAGTAGCTTTTATTAGAGAAAGGATAAA 62

Qy   2268 tgcagtccagatcatggtacttaggcaacagtaccaaggccttctg 2313
           ||||||||  | |||||||| |||||||| || |  | |||||  |
Db     61 CGCAGTCCAAGTTATGGTACTAAGGCAACAATATCGGGTCCTTCAG 16


RESULT   9
AZ814037
LOCUS       AZ814037      632 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0081N06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0081N06 R, DNA sequence.
ACCESSION   AZ814037
VERSION     AZ814037.1  GI:12983945
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0081  row: N  column: 06
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 632.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Mus musculus"
```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0081N06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 203 a 115 c 138 g 176 t
ORIGIN

Query Match 8.6%; Score 253; DB 13; Length 632;
Best Local Similarity 66.6%; Pred. No. 2.4e-44;
Matches 383; Conservative 0; Mismatches 180; Indels 12; Gaps 1;

```
Qy   1737 aagagaacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgt 1796
          ||||||||||||  || | ||  | || ||  |  | ||||| ||     |  | ||| ||
Db      2 AAGAGAACCCGTTACCTTAACTTTGGCAGTTCTATTAGGATTGGGAGTAGTGGCTGGAGT 61

Qy   1797 aggaacagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtga 1856
          ||| ||||| || |||||| | || |     || ||  |    ||  ||
Db     62 AGGTACAGGAACTGCTGCCTTAATTAAGACCCCCCAATACTATGAAGAA---------- 110

Qy   1857 gctacatgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctaga 1916
           |||| ||| || |||    | ||||| || |  || |  ||||| | | |||  ||||
Db    111 -CTACGTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAACCAAATTAGA 169

Qy   1917 agagtccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgct 1976
          ||| ||  | |||||  |||| |||||||| ||||||||  | ||||||||||  |  |
Db    170 AGAATCTTTAACTTCCCTGTCCGAAGTGGTGCTACAGAATAGAAGGGGATTAGACTTATT 229

Qy   1977 gtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtaga 2036
           || || | |||||| ||  | ||||| ||| ||||||||| |||||| || ||||| ||
Db    230 ATTCCTTAAAGAAGGAGGACTCTGTGCTGCCCTAAAAGAAGGATGTTGTTTTTATGTTGA 289

Qy   2037 tcactcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcg 2096
          || |||||||  |||| ||| || |||  ||| |||||| ||    |||| |  |||
Db    290 CCATTCAGGAGTAATCAAAGATTCTATGGCCAAACTTAGAGAACGCCTAGATATACGTAA 349

Qy   2097 aagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggat 2156
          ||| ||||||||     | || || ||||| ||| | ||||| || | ||| ||||||
Db    350 AAGAGAAAGAGAAAGCCAACAAGGATGGTTCGAAAGCTGGTTTAATAAGTCCCCTTGGCT 409

Qy   2157 gaccaccctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagt 2216
           ||||| || || ||  |  |  | || ||  || |    ||  | || || |||||  |
Db    410 CACCACTCTCCTCTCCACCATAGCAGGACCTTTAATTACACTTATGCTTTTGCTTACTTT 469

Qy   2217 tgggccttgcttaattaataggtttgttgccttttgttagagaacgagtgagtgcagtcca 2276
          |||  | ||| |  |||||| ||| || || ||| |||||||| |  | | |||||| ||
Db    470 TGGCTCCTGCATCCTTAATAAGTTAGTCGCTTTTATTAGAGAAAGGATAAATGCAGTACA 529

Qy   2277 gatcatggtacttaggcaacagtaccaaggccttc 2311
          | | |||||||| |||||||| || |  | |||||
Db    530 GGTTATGGTACTAAGGCAACAATATCGGGTCCTTC 564
```

RESULT 10
AZ642336
LOCUS AZ642336 624 bp DNA GSS 14-DEC-2000
DEFINITION 1M0505M18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0505M18 F, DNA sequence.
ACCESSION AZ642336
VERSION AZ642336.1 GI:11768843
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 624)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0505 row: M column: 18
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 624.

FEATURES Location/Qualifiers
source 1. .624
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0505M18"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 203 a 116 c 131 g 174 t
ORIGIN

Query Match 8.3%; Score 246.8; DB 13; Length 624;
Best Local Similarity 66.1%; Pred. No. 5.2e-43;
Matches 378; Conservative 0; Mismatches 182; Indels 12; Gaps 1;

```
Qy   1742 aacccgtatcccttaccctagctgtaatgctcggattagggacggccgttggcgtaggaa 1801
          |||||||  || | ||  | || ||  |  | ||||| ||    || | ||| ||||| |
Db      1 AACCCGTTACCTTAACTTTGGCAGTTCTATTAGGATTGGGAGTAGCAGCTGGAGTAGGTA 60
```

```
Qy   1802 cagggacagctgccctgatcacaggaccacagcagctagagaaaggacttggtgagctac 1861
          |||| || |||||| | || |     || ||  |    ||  ||            ||||
Db     61 CAGGAACCGCTGCCTTAATTAAGACCCCCCAATACTATGAAGAA------------CTAC 108

Qy   1862 atgcggccatgacagaagatctccgagccttaaaggagtctgttagcaacctagaagagt 1921
           ||| || |||      |||||  || |  || |  ||||| | | |||  ||||||| |
Db    109 GTGCAGCTATGGATATTGATCTTAGAACTATAGAACAGTCTATAACCAAATTAGAAGAAT 168

Qy   1922 ccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctgctgtttc 1981
          |  | |||||  |||| |||||||| || || ||  | ||||||||||  |  | || |
Db    169 CTTTAACTTCCCTGTCCGAAGTGGTGCTGCAAAATAGAAGGGGATTAGACTTATTATTCC 228

Qy   1982 taagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcact 2041
          | | ||||||  ||  | ||||| ||||||||||||||||||||| || ||||| || || |
Db    229 TTAAAGAAGGAGGACTCTGTGCTGCCTTAAAAGAAGAATGTTGTTTTTATGTTGACCATT 288

Qy   2042 caggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgtcgaaggg 2101
          ||||||  |||| ||| || |||   || |||||| ||    |||| |  |||  ||| |
Db    289 CAGGAGTAATCAAAGATTCTATGGCTAAACTTAGAGAACGCCTAGATATACGTAAAAGAG 348

Qy   2102 aaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttggatgacca 2161
          |||||||     | || || |||||||| | ||||| || | ||| |||||| | ||||
Db    349 AAAGAGAAAGCCAACAAGGATGGTTTGAAAGCTGGTTTAATAAGTCCCCTTGGCTCACCA 408

Qy   2162 ccctgctttctgctctgacggggcccctagtagtcctgctcctgttacttacagttgggc 2221
          | || || ||  || |  | || ||  || |    ||  | || || |||||  |||| |
Db    409 CTCTCCTCTCCACTATAGCAGGACCTTTAATTACACTTATGCTTTTGCTTACTTTTGGCC 468

Qy   2222 cttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtccagatca 2281
          | ||| |  |||||| ||| || || ||| |||||||| |  | |  |||||||  | |
Db    469 CATGCATCCTTAATAAGTTAGTAGCTTTTATTAGAGAAAGGATAAACGCAGTCCAAGTTA 528

Qy   2282 tggtacttaggcaacagtaccaaggccttctg 2313
          |||| || || ||||| || |  | |||||
Db    529 TGGTCCTAAGACAACAATATCGGGTCCTTCAG 560
```

RESULT 11
AZ850764/c
LOCUS AZ850764 641 bp DNA GSS 21-FEB-2001
DEFINITION 2M0152C18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0152C18 R, DNA sequence.
ACCESSION AZ850764
VERSION AZ850764.1 GI:13036088
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 641)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0152 row: C column: 18
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 641.
FEATURES Location/Qualifiers
source 1. .641
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0152C18"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 197 a 134 c 123 g 187 t
ORIGIN

Query Match 8.3%; Score 245.2; DB 13; Length 641;
Best Local Similarity 63.5%; Pred. No. 1.1e-42;
Matches 414; Conservative 0; Mismatches 223; Indels 15; Gaps 2;

```
Qy   1610 gtgtttccacctcagtcttcaaccaatccaaagatttgtgtgtcatggtccaaatcgtcc 1669
          |||| || |  || || || ||    |||||||||| ||  |  |||| ||  |  | |
Db    641 GTGTGTCTATGTCTGTTTTTAATAGTTCCAAAGATTTCTGCATTTTGGTTCAGCTTATTC 582

Qy   1670 cccgagtgtactaccatcctgaggaagtggtccttgatgaatatgactatcggtataacc 1729
          |  || |    || |||  |||        |  | ||  ||| |||  |||||     ||
Db    581 CTAGACTCCTGTATCATGATGATAGCTCCTTTTTAGACAAATTTGAGCATCGGGT---CC 525

Qy   1730 gaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggattagggacggccg 1789
          |    | ||||||||||||  || | ||  | || ||  |  | ||||| ||    || |
Db    524 GCTGGAGAAGAGAACCCGTTACCTTAACTTTGGCAGTTCTATTAGGATTGGGAGTAGCGG 465

Qy   1790 ttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaaaggac 1849
          ||| ||||| ||||| || |||||| | || |     || |            ||
Db    464 CTGGAGTAGGTACAGGAACCGCTGCCTTAATTAAGACCCCCC------------AATACT 417

Qy   1850 ttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgttagca 1909
          ||  || |||| ||| || |||   |  |||||  || |  || |  ||||| | | ||
Db    416 ATGAAGAACTACGTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAACCA 357

Qy   1910 acctagaagagtccctgacttctttgtctgaagtggttctacagaaccggaggggattag 1969
          |  |||||||  ||  | |||||  |||| |||||| | ||||||||  | ||||||||||
Db    356 AATTAGAAGAATCTTTAACTTCCCTGTCCGAAGTGTTGCTACAGAATAGAAGGGGATTAG 297

Qy   1970 atctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttct 2029
          |  |  | || || | |||||| ||  | |||||  ||| |||||||  |||||||  || |
Db    296 ACTTATTATTCCTTAAAGAAGGAGGACTCTGTGCTGCCCTAAAAGAGGAATGTTGTTTTT 237

Qy   2030 atgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagttagaga 2089
          |||| || || |||||||  |||| ||| || |||  ||| |||||| ||    |||| |
Db    236 ATGTTGACCATTCAGGAGTAATCAAAGATTCTATGGCCAAACTTAGAGAACGCCTAGATA 177

Qy   2090 ggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctc 2149
            | |  ||| |||||||||     | || || ||||| ||| | |||||| || | ||  |
Db    176 TACTTAAAAGAGAAAGAGAAAGCCAACAAGGATGGTTCGAAAGCTGGTTTAATAAGTTCC 117

Qy   2150 cttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctcctgttac 2209
          ||||| | ||||| || || ||  |  |  | || ||  || |    ||  | || || |
Db    116 CTTGGCTCACCACTCTCCTCTCCACCATAGCAGGACCTTTAATTACACTTATGCTTTTGC 57
```

```
Qy   2210 ttacagttgggccttgcttaattaataggtttgttgcctttgttagagaacg 2261
          ||||  |||| || ||| |  |||||| ||| || || ||| |||||||| |
Db     56 TTACTTTTGGCCCCTGCATCCTTAATAAGTTGGTAGCTTTTATTAGAGAAAG 5


RESULT  12
AZ349077/c
LOCUS       AZ349077      493 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0085K21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0085K21 R, DNA sequence.
ACCESSION   AZ349077
VERSION     AZ349077.1  GI:10428314
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0085  row: K  column: 21
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0085K21"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      141 a    103 c     89 g    160 t
ORIGIN


  Query Match             8.2%;  Score 242.6;  DB 13;  Length 493;
  Best Local Similarity   70.7%;  Pred. No. 4.4e-42;

  Matches  323;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy   1855 gagctacatgcggccatgacagaagatctccgagccttaaaggagtctgttagcaaccta 1914
           || |||| ||| || |||   |  |||||  || |  || |  |||||  | | |||  ||
Db    491 GAACTACGTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAACCAAGTTA 432

Qy   1915 gaagagtccctgacttctttgtctgaagtggttctacagaaccggaggggattagatctg 1974
           |||||  ||  | |||||  |||||  |||||||| ||||||||  | ||||||||||  |
Db    431 GAAGAATCTTTAACTTCCTTGTCCGAAGTGGTGCTACAGAATAGAAGGGGATTAGACTTA 372

Qy   1975 ctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgcttctatgta 2034
           | |||||  | ||||||  ||   | |||||  ||| ||||||||||||||||  || |||||
Db    371 TTATTTCTTAAAGAAGGAGGACTCTGTGCTGCCCTAAAAGAAGAATGTTGTTTTTATGTT 312

Qy   2035 gatcactcaggagccatcagagactccatgaacaagcttagaaaaaagttagagaggcgt 2094
           ||  || |||||||  ||||  |||  ||  |||  |||  ||||||  ||     ||||  |  |||
Db    311 GACCATTCAGGAGTAATCAAAGATTCTATGGCCAAACTTAGAGAACGCCTAGATATACGT 252

Qy   2095 cgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtctccttgg 2154
           |||  ||||||||      | ||  ||  |||||  |||  |  |||||  ||  |  |||  ||||||
Db    251 AAAAGAGAAAGAGAAAGCCAACAAGGATGGTTCGAAAGCTGGTTTAATAAGTCCCCTTGG 192

Qy   2155 atgaccaccctgctttctgctctgacggggcccctagtagtcctgctcctgttacttaca 2214
           |  |||||  ||  ||  ||   |  |  ||  ||  ||  ||  |     ||  |  ||  ||  |||||
Db    191 CTCACCACTCTCCTCTCCACCATAACAGGACCTTTAATTACACTTATGCTTTTGCTTACT 132

Qy   2215 gttgggccttgcttaattaataggtttgttgcctttgttagagaacgagtgagtgcagtc 2274
           ||||  ||  |||  |  ||||||  |||  ||  ||  |||  ||||||||  |   |  |  ||||||
Db    131 TTTGGCCCCTGCATCCTTAATAAGTTGGTAGCTTTTATTAGAGAAAGGATAAATGCAGTA 72

Qy   2275 cagatcatggtacttaggcaacagtaccaaggccttc 2311
           |||  |  ||||||||  ||||||||  ||  |   |  |||||
Db     71 CAGGTTATGGTACTAAGGCAACAATATCGGGTCCTTC 35


RESULT  13
AZ620796/c
LOCUS       AZ620796      650 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0453I12R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0453I12 R, DNA sequence.
ACCESSION   AZ620796
VERSION     AZ620796.1  GI:11742986
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 650)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0453  row: I  column: 12
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 650.
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
```

```
                     /clone="UUGC1M0453I12"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      180 a    131 c    128 g    211 t
ORIGIN

  Query Match             8.1%;  Score 240.8;  DB 13;  Length 650;
  Best Local Similarity   67.0%;  Pred. No. 1e-41;
  Matches  363;  Conservative    0;  Mismatches  167;  Indels   12;  Gaps    1;

Qy   1772 tcggattagggacggccgttggcgtaggaacagggacagctgccctgatcacaggaccac 1831
           | ||||| ||    || | ||| ||||| ||||| |  |||||| | || |     || |
Db    649 TAGGATTGGGAGTAGCAGCTGGAGTAGGTACAGGAATCGCTGCCTTAATTAAGACCCCCC 590

Qy   1832 agcagctagagaaaggacttggtgagctacatgcggccatgacagaagatctccgagcct 1891
             | ||| |             || |||| ||| || |||      |||||  || |
Db    589 CAAATCTATA------------AGAACTACGTGCAGCTATGGATATTGATCTTAGAACTA 542

Qy   1892 taaaggagtctgttagcaacctagaagagtccctgacttctttgtctgaagtggttctac 1951
           || |  ||||| | | |||  ||||||| ||  | |||||  |||| |||||||| || |
Db    541 TAGAACAGTCTATAACCAAATTAGAAGAATCTTTAACTTCCCTGTCCGAAGTGGTGCTGC 482

Qy   1952 agaaccggaggggattagatctgctgtttctaagagaaggtgggttatgtgcagccttaa 2011
           | ||  | ||||||||||  |  | || || | |||||| ||  | |||| |||||||
Db    481 AAAATAGAAGGGGATTAGACTTATTATTCCTTAAAGAAGGAGGACTCTGTGCTGCCTTAA 422

Qy   2012 aagaagaatgttgcttctatgtagatcactcaggagccatcagagactccatgaacaagc 2071
           |||||||||||||| || |||||| || || |||||||  |||| ||| || |||  ||| |
Db    421 AAGAAGAATGTTGTTTTTATGTTGACCATTCAGGAGTAATCAAAGATTCTATGGCCAAAC 362

Qy   2072 ttagaaaaaagttagagaggcgtcgaagggaaagagaggctgaccaggggtggtttgaag 2131
           ||||| ||    |||| |  |||  ||| |||||||||     | || || |||||||||
Db    361 TTAGAGAACGCCTAGATATACGTAAAAGAGAAAGAGAAAGCCAACAAGGATGGTTTGAAA 302

Qy   2132 gatggttcaacaggtctccttggatgaccaccctgctttctgctctgacggggcccctag 2191
           | ||||| || | ||| |||||| | ||||| || || ||  || |  | || ||  ||
Db    301 GCTGGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACTATAGCAGGACCTTTAA 242

Qy   2192 tagtcctgctcctgttacttacagttgggccttgcttaattaataggtttgttgcctttg 2251
           |    ||  | || || ||||| |||| || ||| |  |||||| ||| || || |||
Db    241 TTACACTTATGCTTTTGCTTACTTTTGGCCCATGCATCCTTAATAAGTTAGTAGCTTTTA 182

Qy   2252 ttagagaacgagtgagtgcagtccagatcatggtacttaggcaacagtaccaaggccttc 2311
           |||||||| |  | |  |||||||  | |||||||| |||||||| || |  | |||||
Db    181 TTAGAGAAAGGATAAACGCAGTCCAAGTTATGGTACTAAGGCAACAATATCGGGTCCTTC 122

Qy   2312 tg 2313
             |
Db    121 AG 120


RESULT  14
```

```
AZ813042/c
LOCUS       AZ813042      625 bp    DNA              GSS       20-FEB-2001
DEFINITION  2M0080E11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0080E11 F, DNA sequence.
ACCESSION   AZ813042
VERSION     AZ813042.1  GI:12982881
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000    Std Error: 0.00
            Plate: 0080  row: E  column: 11
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 625.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0080E11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      191 a    128 c    120 g    186 t
ORIGIN

  Query Match             8.1%;  Score 240.6;  DB 13;  Length 625;
  Best Local Similarity   63.8%;  Pred. No. 1.1e-41;
  Matches  404;  Conservative    0;  Mismatches  214;  Indels   15;  Gaps    2;

Qy   1574 acaggtggtgggcatgcaatactgggttaacccccctgtgtttccacctcagtcttcaacc 1633
           |||    ||| || |||||||  ||  | || || ||||| || |  || || || ||
Db    618 ACACAGTGTGAGCTTGCAATAACGGTCTCACTCCTTGTGTGTCTATGTCTGTTTTTAATA 559

Qy   1634 aatccaaagatttgtgtgtcatggtccaaatcgtcccccgagtgtactaccatcctgagg 1693
```

```
Db    558 GTTCTAAAGATTTCTGCATTTTGGTTCAGCTTATTCCCAGACTCTTGTATCATGATGATA 499

Qy   1694 aagtggtccttgatgaatatgactatcggtataaccgaccaaaaagagaacccgtatccc 1753
Db    498 GCTCTTTTTTAGATAAATTTGAACATCGG---GTCCGCTGGAAAAGAGAACCCGTTACCT 442

Qy   1754 ttaccctagctgtaatgctcggattagggacggccgttggcgtaggaacagggacagctg 1813
Db    441 TAACTTTGGCAGTTCTATTAGGATTGGGAGTAGCAGCTGGAGTAGGTACAGGAACCGCTG 382

Qy   1814 ccctgatcacaggaccacagcagctagagaaaggacttggtgagctacatgcggccatga 1873
Db    381 CCTTAATTAAGACCCCCC----------AATACTATGAAGAACTACGTGCAGCTATGG 334

Qy   1874 cagaagatctccgagccttaaaggagtctgttagcaacctagaagagtccctgacttctt 1933
Db    333 ATATTGATCTTAGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAACTTCCC 274

Qy   1934 tgtctgaagtggttctacagaaccggaggggattagatctgctgtttctaagagaaggtg 1993
Db    273 TGTCTGAAGTGGTGCTGCAAAATAGAAGGGGATTAGACTTATTATTCCTTAAAGAAGGAG 214

Qy   1994 ggttatgtgcagccttaaaagaagaatgttgcttctatgtagatcactcaggagccatca 2053
Db    213 GACTCTGTGCTGCCTTAAAAGAAGAATGTTGTTTTTATGTTGACCATTCAGGAGTAATCA 154

Qy   2054 gagactccatgaacaagcttagaaaaaagttagagaggcgtcgaagggaaagagaggctg 2113
Db    153 AAGATTCTATGGCTAAACTTAGAGAACGCTTAGATATACGTAAAAGAGAAAGAGAGAGCC 94

Qy   2114 accaggggtggtttgaaggatggttcaacaggtctccttggatgaccaccctgcttttctg 2173
Db     93 AACAAGGATGGTTCGAAAGCTGGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCA 34

Qy   2174 ctctgacggggcccctagtagtcctgctcctgt 2206
Db     33 CTATAGCAGGACCTTTAATTACACTTATGCTTT 1
```

RESULT 15
AZ836232

```
LOCUS       AZ836232      592 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0131E01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0131E01 F, DNA sequence.
ACCESSION   AZ836232
VERSION     AZ836232.1  GI:13006140
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0131  row: E  column: 01
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 592.
```

```
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0131E01"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      188 a    112 c    127 g    165 t
ORIGIN
```

Query Match 7.9%; Score 234.6; DB 13; Length 592;
Best Local Similarity 65.6%; Pred. No. 2.2e-40;
Matches 363; Conservative 0; Mismatches 179; Indels 11; Gaps 1;

```
Qy   1728 ccgaccaaaaagagaacccgtatcccttaccctagctgtaatgctcggattagggacggc 1787
Db     33 CCGCTGGAAAAGAGAACCCGTTACCTTAACTTTGGCAGTTCTATTAGGATTGGGAGTAGC 92

Qy   1788 cgttggcgtaggaacagggacagctgccctgatcacaggaccacagcagctagagaaagg 1847
Db     93 AGCTGGAGTAGGTACAGGAACTGCTGCCTTAATTAAGACCCCCCCAATACTA-------- 144

Qy   1848 acttggtgagctacatgcggccatgacagaagatctccgagccttaaaggagtctgttag 1907
Db    145 ---TGAAGAACTACGTGCAGCTATGGATATTGATCTTAGAACTATAGAACAGTCTATAAC 201

Qy   1908 caacctagaagagtccctgacttctttgtctgaagtggttctacagaaccggaggggatt 1967
Db    202 CAAATTAGAAGAATCTTTAACTTCCCTGTCCAAAGTGGTGCTGCAAAATAGAAGGGGATT 261

Qy   1968 agatctgctgtttctaagagaaggtgggttatgtgcagccttaaaagaagaatgttgctt 2027
Db    262 AGACTTATTATTCCTTAAAGAAGGAGGACTCTGTGCTGCCTTAAAAGAGGAATGTCGTTT 321

Qy   2028 ctatgtagatcactcaggagccatcagagactccatgaacaagcttagaaaaaagttaga 2087
Db    322 TTATGTTGACCATTCAGGAGTAATCAAAGATTCTATGGCTAAACTTAGAGAACGCCTAGA 381

Qy   2088 gaggcgtcgaagggaaagagaggctgaccaggggtggtttgaaggatggttcaacaggtc 2147
Db    382 TATACGTAAAAGAGAAAGAGAAAGCCAACAAGGATGGTTTGAAAGCTGGTTTAATAAGTC 441

Qy   2148 tccttggatgaccaccctgctttctgctctgacggggcccctagtagtcctgctcctgtt 2207
Db    442 CCCTTGGCTCACCACTCTCCTCTCCACTATAGCAGGACCTTTAATTACACTTATGCTTTT 501

Qy   2208 acttacagttgggccttgcttaattaataggtttgttgccttttgttagagaacgagtgag 2267
Db    502 GCTTACTTTTGGCCCATGTATCCTTAATAAGTTAGTAGCTTTTATTAGAGAAAGGATAAA 561

Qy   2268 tgcagtccagatc 2280
```

```
          |||||||||  ||
Db    562 TGCAGTCCAAGTC 574
```

Search completed: February 23, 2002, 23:37:18
Job time: 9439 sec

```
OM of: US-09-171-553B-10 to: EST:*   out_format : pfs

Date: Feb 24, 2002  8:32 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09171553/runat_22022002_122825_11137/app_query.fasta_1.3312
 -DB=EST -QFMT=fastap -SUFFIX=feb22std.rst -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09171553_@CGN1_1_9008 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-171-553B-10
Query length: 657
Database: EST:*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 3515.700000


score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est1:AW657531     +  888.00 1216.45  1.4e-58   549  ! AW657531 110922 MARC 1BOV Bos t
gb_gss:AZ970955      -  875.50 1196.21  1.8e-57   754  ! AZ970955 2M0244H16F Mouse 10kb
gb_gss:BH059027      +  859.50 1172.69  3.7e-56   882  ! BH059027 RPCI-24-326G3.TJ RPCI-
gb_gss:BH122287      +  843.50 1151.93  5.3e-55   774  ! BH122287 RPCI-24-288P13.TJ RPCI
gb_gss:AZ351242      -  814.00 1112.30  8.6e-53   701  ! AZ351242 1M0089L09F Mouse 10kb
gb_est2:BI151181     +  787.00 1073.36  1.3e-50   845  ! BI151181 602916935F1 NCI_CGAP_L
gb_gss:AZ850764      -  770.50 1053.32  1.7e-49   641  ! AZ850764 2M0152C18R Mouse 10kb
gb_gss:AZ813042      -  757.50 1035.67  1.6e-48   625  ! AZ813042 2M0080E11F Mouse 10kb
gb_est2:BI106669     +  757.00 1031.27  2.8e-48   921  ! BI106669 602893272F1 NCI_CGAP_L
gb_gss:AZ356870      -  734.50 1003.74  9.6e-47   644  ! AZ356870 1M0098F09F Mouse 10kb
gb_gss:AZ376517      -  730.50  998.76  1.8e-46   610  ! AZ376517 1M0130J20F Mouse 10kb
gb_gss:AZ430249      -  729.00  997.00  2.3e-46   591  ! AZ430249 1M0214G11R Mouse 10kb
gb_gss:AZ642336      +  721.50  986.16  9.1e-46   624  ! AZ642336 1M0505M18F Mouse 10kb
gb_gss:AZ337993      +  721.50  985.44  1.0e-45   673  ! AZ337993 1M0068F24R Mouse 10kb
gb_gss:AZ814037      +  713.50  975.04  3.8e-45   632  ! AZ814037 2M0081N06R Mouse 10kb
gb_gss:AZ741412      +  713.50  974.63  4.0e-45   659  ! AZ741412 RPCI-24-74C13.TJ RPCI-
gb_gss:AZ327473      -  711.00  971.07  6.3e-45   668  ! AZ327473 1M0050K14R Mouse 10kb
gb_gss:AZ375836      -  696.50  952.72  6.7e-44   565  ! AZ375836 1M0129J20F Mouse 10kb
gb_gss:AZ499035      +  693.00  948.45  1.2e-43   534  ! AZ499035 1M0336P05R Mouse 10kb
gb_gss:AZ739547      -  693.00  946.19  1.5e-43   676  ! AZ739547 RPCI-24-73L17.TJ RPCI-
gb_gss:AZ417785      -  685.50  937.29  4.8e-43   583  ! AZ417785 1M0193B08R Mouse 10kb
gb_est2:BI109841     +  681.50  928.91  1.4e-42   787  ! BI109841 602901065F1 NCI_CGAP_M
gb_gss:AZ620796      -  671.50  916.99  6.5e-42   650  ! AZ620796 1M0453I12R Mouse 10kb
gb_gss:AZ836232      +  670.00  915.82  7.6e-42   592  ! AZ836232 2M0131E01F Mouse 10kb
gb_est2:BF608155     +  670.00  912.54  1.2e-41   833  ! BF608155 MY1_001169 Mouse 9-day
gb_est1:BE569478     +  669.00  910.48  1.5e-41   895  ! BE569478 601328494F1 NCI_CGAP_M
gb_gss:AZ416306      -  662.50  905.29  2.9e-41   605  ! AZ416306 1M0191D01R Mouse 10kb
gb_est2:BF159914     +  659.00  899.36  6.2e-41   680  ! BF159914 601767582F1 NCI_CGAP_L
gb_gss:BH054319      -  653.00  890.50  1.9e-40   724  ! BH054319 RPCI-24-398E11.TV RPCI
gb_gss:AZ336017      -  652.50  891.22  1.8e-40   625  ! AZ336017 1M0066P05F Mouse 10kb
gb_est2:BF181106     +  645.00  876.70  1.1e-39   969  ! BF181106 601805164F1 NCI_CGAP_M
gb_gss:AZ764389      +  642.50  877.72  1.0e-39   609  ! AZ764389 1M0560M15F Mouse 10kb
gb_gss:AZ647666      +  638.00  870.48  2.5e-39   679  ! AZ647666 1M0514J12F Mouse 10kb
gb_est2:BI106591     +  626.50  853.99  2.1e-38   728  ! BI106591 602893174F1 NCI_CGAP_L
gb_est2:BI158931     +  625.50  852.78  2.5e-38   716  ! BI158931 602921838F1 NIH_CGAP_M
gb_gss:AZ989131      +  624.50  852.24  2.6e-38   656  ! AZ989131 2M0272E22F Mouse 10kb
gb_gss:AZ349077      -  619.00  847.41  4.9e-38   493  ! AZ349077 1M0085K21R Mouse 10kb
gb_est2:BI082444     +  618.00  842.01  9.8e-38   750  ! BI082444 602877669F1 NCI_CGAP_M
gb_gss:AZ960032      -  612.00  835.22  2.3e-37   644  ! AZ960032 2M0227K23R Mouse 10kb
gb_gss:AZ397096      -  611.50  834.08  2.7e-37   675  ! AZ397096 1M0161J22R Mouse 10kb
gb_gss:AZ996293      -  603.50  824.77  8.9e-37   566  ! AZ996293 2M0282L19F Mouse
gb_est2:BF786434     +  602.00  819.37  1.8e-36   801  ! BF786434 602112966F1 NCI_C
gb_gss:AZ887402      -  599.00  818.42  2.0e-36   575  ! AZ887402 RPCI-24-176N10.TJ
gb_gss:BH055839      +  596.00  813.57  3.7e-36   620  ! BH055839 RPCI-24-93P16.TVB
gb_gss:AZ637496      -  596.00  813.08  4.0e-36   653  ! AZ637496 1M0496I23R Mouse

seq_name: gb_est1:AW657531

seq_documentation_block:
LOCUS       AW657531      549 bp    mRNA            EST       25-APR-2001
DEFINITION  110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW657531
VERSION     AW657531.1  GI:7423429
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 94  row: E  column: 12
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT      145 a    144 c    121 g    139 t
ORIGIN


alignment_scores:
            Quality:  888.00            Length:    183
              Ratio:  5.074               Gaps:      1
 Percent Similarity:  95.628   Percent Identity: 89.617

alignment_block:
 US-09-171-553B-10 x AW657531   ..

 Align seg 1/1  to: AW657531  from: 1  to: 549

     290 SerProGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGl 306
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 TCCCCCGGTGTTCCTGTTAAGACAGGACAGAGACTCTTCAGTCTCATCCA 52

     306 nGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerC 323
         ||||||||||||||||||||||||||||||||||||||:::|||||||
```

```
     53 GGGAGCTTTCCAAGCCATCAACTCCACCGACCCCGATGCCAGTTCTTCTT 102

    323 ysTrpLeuCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLys 339
        ||||||||||||||||||||||||||||||||||||||||||||||||:::
    103 GTTGGCTTTGTCTATCCTCAGGGCCTCCTTATTATGAGGGAATGGCTAGA 152

    340 GluArgLysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGl 356
        |||   ||||||||||||||||||||||||||||||||||||||||||||
    153 GAAGGGAAATTCAATGTGACCAAAGAGCATAGAAATCAATGTACATGGGG 202

    356 ySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysI 373
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    203 GTCCCGAAATAAGCTCACCCTCACTGAAGTTTCCGGGAAGGGGACATGCA 252

    373 leGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValVal 389
        ||||||||||||||||||||||||||||||||||   ||||||:::|||
    253 TAGGAAAAGCTCCCCCATCCCACCAACACCTTTGCAATAGTACTATGGTT 302

    390 TyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTr 406
        ||||||||||||   |||||||||||||||||||||||||||||   ||
    303 TATGAGCAGGCCTTAGAAAATCAGTATTTAGTACCTGGTTATAACTGATG 352

    406 pTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheA 423
        ||||||||||||||||||||||||||||||||||||||||:::|||||||
    353 GTGGGCATGCAATACTGGGTTAACTCCCTGTGTTTCTACCACAGTCTTTA 402

    423 snGlnSerLysAspLeuCysValMetValGlnIleValProArgValTyr 439
        ||||||||||||||:::|||||||||||||||||:::|||||||   |||:::
    403 ACCAATCCAAAGATTTCTGTGTCATGGTCCAGCTCGTCCCCTGGGTACAC 452

    440 TyrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnAr 456
        ||||||||||||||||||||:::|||||||||||||||||||   :::||
    453 TACCATCCTGAGGAAGTAGTCATCGATGAGTATGACTATCGGCCTACCCG 502

    456 gProLysArgGluProValSer.LeuThrLeuAlaValMetLeuGly 471
        |   ||||||||||||||::: ||||||||||||||||:::|||||
    503 CTCAAAAAGAGAGCCCGTGACCCTTAACCCTAGCCGTTATACTAGGA 549

seq_name: gb_gss:AZ970955

seq_documentation_block:
LOCUS       AZ970955      754 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0244H16F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0244H16 F, DNA sequence.
ACCESSION   AZ970955
VERSION     AZ970955.1  GI:13842182
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 754)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0244  row: H  column: 16
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 754.
FEATURES             Location/Qualifiers
     source          1. .754
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0244H16"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      226 a    160 c    150 g    218 t
ORIGIN


alignment_scores:
           Quality:  875.50              Length:    244
             Ratio:  3.998                 Gaps:      3
 Percent Similarity:  89.754   Percent Identity: 67.213

alignment_block:
 US-09-171-553B-10 x AZ970955/rev   ..

 Align seg 1/1  to reverse of: AZ970955  from: 1  to: 754

    407 TrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAs 423
        ||||||||||||||||||||   ||||||||||||   |||   |||||
    751 TGGGCTTGCAATACCGGTCTCATTCCTTGTGTGTCTATGTCTGATTTTAA 702

    423 nGlnSerLysAspLeuCysValMetValGlnIleValProArgValTyrT 440
        |:::|||||||||:::|||:::::|||||:::::|||||:::   |
    701 TAGTTCCAAAGATTTCTGCATTTTGGTTCAGCTTATTCCTAGACTCCTGT 652

    440 yrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArg 456
        |||||   ::::::      ||||||:::::::::::::|||      |||
    651 ATCATGATGATAGCTCCTTTTTAGACAAATTTGAGCATCGGGTC...CGC 605

    457 ProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGl 473
           :::|||||||||:::::|||||||||:::|||:::||||||||||
    604 TGGAGAAGAGAGCCCATTACCTTAACTTTGACAGTTCTATTAGGATTAGG 555

    473 yThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProG 490
        |:::|||:::|||||||||||||||||||||||||||||      ||||
    554 AGTAGCGGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAAGACCCCCC 505

    490 ln.GlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAs 506
        ||       |||            |||||:::|||||||||      ||
    504 AATACTAATGAA............GAACTACGTGCAGCTATGGATGTTGA 467

    506 pLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrS 523
        |||||||:::::::::::::|||:::::::::|||||||||||||||||||
    466 TCTTAGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAACTT 417

    523 erLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPhe 539
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    416 CCCTGTCCGAAGTGGTGCTACAGAATAGAAGGGGATTAGACTTATTATTC 367
```

```
     540 LeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTy 556
     366 CTTAAAGAAGGAGGACTCTGTGCTGCCCTAAAAGAAGAATGTTGTTTTTA 317

     556 rValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysL 573
     316 TGTTGACCATTCAGGAGTAATCAAAGATTCTATGGCCAAACTTAGAGAAC 267

     573 ysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGlu 589
     266 GCCTAGATACACGTAAAAGGGAAAGAGAAAGCCAACAAGGATGGTTCGAA 217

     590 GlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuTh 606
     216 AGCTGGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACCATAGC 167

     606 rGlyProLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuI 623
     166 AGGACCTTTGATTACTCTTATGCTTTTGCTTACTTTTGGCCCCTGCATCC 117

     623 leAsnArgPheValAlaPheValArgGluArgValSerAlaValGlnIle 639
     116 TTAATAAGTTAGTAGCTTTTATTAGAGAAAGGATAAATGCAGTACAGGTT 67

     640 MetValLeuArgGlnGlnTyrGlnGlyLeu 649
      66 ATGGTACTAAGGCAACAATATCGGGTCCTT 37

seq_name: gb_gss:BH059027

seq_documentation_block:
LOCUS       BH059027      882 bp    DNA             GSS       18-JUL-2001
DEFINITION  RPCI-24-326G3.TJ RPCI-24 Mus musculus genomic clone RPCI-24-326G3,
            DNA sequence.
ACCESSION   BH059027
VERSION     BH059027.1  GI:14868646
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-326G3.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 326  row: G  column: 3
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .882
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-326G3"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      248 a    183 c    201 g    250 t
ORIGIN


alignment_scores:
            Quality:  859.50              Length:    309
              Ratio:  3.627                 Gaps:      6
 Percent Similarity:  76.699   Percent Identity: 57.929

alignment_block:
 US-09-171-553B-10 x BH059027   ..

 Align seg 1/1  to: BH059027  from: 1  to: 882

     298 GlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSe 314
       2 GGGCAAAGACTGTTCAACCTGATTAGAGGAGCCTTTTATGCCCTTAACAG 51

     314 rThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyP 331
      52 CTCGGATCCAGATGCCACTACATACTGCTGGTTATTCCTATCATCAGGCC 101

     331 roProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLys 347
     102 CTCCCTACTCTGAAGGGGTTGCTTTTAATGAAGATTTTAATAAAACAAGT 151

     348 GluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeu.. 363
     152 AGTCATGCC...TTATGCTCATGGGGAACAGGAAAGAAATTAACCCTGTA 198

     364 ....ThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProProS 379
     199 CTGAAGTGTCAGGAGCAGCCCTTGTCTCTGTACAGGGACACCTCCTCCTG 248

     379 erHisGlnHisLeuCysTyrSerThr.ValValTyrGluGlnAlaSerGl 395
     249 CCCACCAACACCTATGTGGTCGGACTTAATTCTGTATCCAGGACAAATAC 298

     395 uAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrG 412
     299 CAACTATTACCTTGTGCCATATCCAGTTGGTTGGTGGGCCTGTAACACAG 348

     412 lyLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspLeu 428
     349 GGCTCACCCCATGTGTATCAACTAGTGTCTTTGACCCCTCTGGTGATTTC 398

     429 CysValMetValGlnIleValProArgValTyrTyrHisProGluGluVa 445
     399 TGTATCATGGTTCATCTTCTATCTCGCATGTATTATTATCCTGCCTCCCG 448

     445 lValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProV 462
     449 CCTGGAAGATGAATATATCAGTAAG......AAGTTGAAAAGAGATCCAG 492

     462 alSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyVal 478
     493 TCTCCATTACTCTAGCTATGTTAATGGGGGTCGGTCTTACAGA.GGAGTA 541

     479 GlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGl 495
     542 GGAATGGGTTTAGCAGCCTTGGTTGAGGGTAGACAAGGAATTCACACCTT 591

     495 yLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysG 512
     592 CTTAAGAGAA.........GCTATAAATGAAGATTTAAGCATGCTACAAA 632

     512 luSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValVal 528
     633 AATCTGTTGATGCATTAGAGAAATCCTTGACGTCCCTGTCAGAAGTAGTA 682
```

```
      529 LeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLe 545
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      683 TTACAAAACAGAAGAGGTTTAGATTTATTGTTTTTAAGGGAAGGTGGACT 732

      545 uCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyA 562
          ||||||||||||||||||||||||||||||||||:::|||||||:::|||
      733 ATGTGCAGCCTTAAAAGAAGAATGCTGCTTCTATGCAGACCACACGGGAA 782

      562 laIleArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArg 578
            :::|||||||||||||:::|||||||||:::::||||||||||:::
      783 TAGTCAGAGATTCTATGCAGAAATTGAGGGAAAGACTGGAAAGGGAA... 829

      579 ArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSe 595
                                  ||||||:::   |||:::::   ||
      830 ........................TGGTTCAAGGTGTGGTACTCCTAATC 855

      595 rProTrpMetThrThrLeuLeuSer 603
          |||||||||||||||||||||||||
      856 TCCATGGATGACCACTTTACTTTCT 880

seq_name: gb_gss:BH122287

seq_documentation_block:
LOCUS       BH122287      774 bp    DNA             GSS       19-JUL-2001
DEFINITION  RPCI-24-288P13.TJ RPCI-24 Mus musculus genomic clone RPCI-24-288P13
            , DNA sequence.
ACCESSION   BH122287
VERSION     BH122287.1  GI:14965799
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-288P13.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 288  row: P  column: 13
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-288P13"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      221 a    162 c    172 g    219 t
ORIGIN
```

```
alignment_scores:
           Quality:  843.50              Length:    262
             Ratio:  3.749                 Gaps:      2
 Percent Similarity:  85.878   Percent Identity: 61.832

alignment_block:
 US-09-171-553B-10 x BH122287   ..

 Align seg 1/1  to: BH122287  from: 1  to: 774

     355 TrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyTh 371
         :::|||::::::::::|||||||||||||:::   |||||||||:::|||
           2 TTTGGGGAGAAAACAAAGTTGACTCTGGCAGCAGTTTCAGGAAGAGGGCT 51

     371 rCysIleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrV 388
          |||:::|||:::::|||   :::      |||||||||   :::|||
      52 TTGTTTGGGCCAGGTACCTCAGGATAAAGGGCACCTCTGTAATCAGACCC 101

     388 alValTyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsn 404
               :::::::::::::::::|||||||||||||||      :::
     102 AGAACATCCAGTCTAGCAAAAGTGGTCAGTATCTAGTGCCCCCCCTAGAC 151

     405 ArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerVa 421
              |||||||||||||||||||||||||||||||||||   |||||
     152 ACAGTGTGGGCTTGCAATACCGGTCTCACTCCTTGTGTGTCTATGTCTGT 201

     421 lPheAsnGlnSerLysAspLeuCysValMetValGlnIleValProArgV 438
         |||||||:::|||||||||:::|||:::::::||||||:::::|||||:
     202 TTTTAATAGTTCCAAAGATTTCTGCATTTTGGTTCAGCTTATTCCTAGAC 251

     438 alTyrTyrHisProGluGluValValLeuAspGluTyrAspTyrArgTyr 454
         ::   ||||||   :::::::::   ||||||:::::::::::::|||
     252 TCCTGTATCATGATGATAGCGCCTTTTTAGACAAATTTGAGCATCGGGTC 301

     455 AsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGl 471
            |||   :::|||||||||||||:::|||||||||||||||:::|||||
     302 ...CGCTGGAGAAGAGAGCCCGTTACCTTAACTTTGGCAGTTCTATTAGG 348

     471 yLeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrG 488
         |||||||:::|||:::|||||||||||||||||||||||||||||||||
     349 ATTAGGAGTAGCGGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAAGA 398

     488 lyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThr 504
           ||||||      |||            ||||||:::|||||||||
     399 CCCCCCAATACTATGAA............GAACTACGTGCAGCTATGGAT 436

     505 GluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLe 521
            |||||||||:::::::::::||:::::::::|||||||||||||
     437 GTTGATCTTAGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTT 486

     521 uThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuL 538
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     487 AACTTCCCTGTCCGAAGTGGTGCTACAAAATAGAAGGGGATTAGACTTAT 536

     538 euPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCys 554
         ||||||||:::|||   |||||||||||||||||||||||||||||||
     537 TATTCCTTAAAGAACGAGGACTCTGTGCTGCCCTAAAAGAAGAATGTTGA 586

     555 PheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuAr 571
         ||||||||||||||||||   :::|||:::|||||||||   ||||||||
     587 TTTTATGTTGACCATTCATGAGTAATCAAAGATTCTATGGCCAAACTTAG 636

     571 gLysLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpP 588
         |:::::|||:::   |||:::||||||   |||:::::|||||||||||
     637 AGAACGCCTAGATATACGTGAAAGAGAAATAGAAAGCCAACAAGGATGGT 686

     588 heGluGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAla 604
         |||||:::|||   |||:::|||||||||:::|||||||||||||||:::
     687 TCGAAAGCTGGGTTAATAAGTCCCCTTGGCTCACCACTCTTCTCTCCACC 736

     605 LeuThrGlyProLeuValValLeuLeuLeuLeuLeu 616
         ::::::|||||||||:::::::|||:::|||||||||
```

```
    737 ATAGCAGGACCTTTGATTACTCTTATGCTTTTGCTT 772

seq_name: gb_gss:AZ351242

seq_documentation_block:
LOCUS       AZ351242      701 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0089L09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0089L09 F, DNA sequence.
ACCESSION   AZ351242
VERSION     AZ351242.1  GI:10430479
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0089  row: L  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 701.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0089L09"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      211 a    149 c    134 g    206 t      1 others
ORIGIN


alignment_scores:
            Quality:  814.00            Length:    234
              Ratio:  3.895               Gaps:      4
 Percent Similarity:  89.316   Percent Identity: 67.521

alignment_block:
 US-09-171-553B-10 x AZ351242/rev   ..

 Align seg 1/1  to reverse of: AZ351242  from: 1  to: 701

     417 ValSerThrSerValPheAsnGlnSerLysAspLeuCysValMetValGl 433
         ||||||   ||||||||||||:::||||||||:::|||:::::||||
     700 GTGTCTATGTCTGTTTTTAATAGTTCCAAAGATTTCTGCATTTTGGTTCA 651

     433 nIleValProArgValTyrTyrHisProGluGluValValLeuAspGluT 450
         |:::::|||||:::   ||||||  ::::::      ||||||   :
     650 ACTTATTCCTAGACTCCTGTATCATGATGATAGCTCCTTTTTAGACANAT 601

     450 yrAspTyrArgTyrAsnArgProLysArgGluProValSer.LeuThrLe 466
         ::::::::|||      |||   :::|||||||||||||::: ||||||||
     600 TTGAGCATCGGGTC...CGCTGGAGAAGAGAACCCGTTACCCTTAACTTT 554

     466 uAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGlyThrA 483
         |||||||:::|||||||||||:::|||||||||||||||||||||||||
     553 GGCAGTTCTATTAGGATTGGGAGTAGCGGTTGGAGTAGGTACAGGAACTG 504

     483 laAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeu 499
         ||||||||||      ||||||      |||           ||||||
     503 CTGCCTTAATTAAGACCCCCCAATACTATGAA.....GAACTA 466

     500 HisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAs 516
         :::|||||||||      |||||||||:::::::::::::|||:::::::
     465 CGTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAACCAA 416

     516 nLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgA 533
         :|||||||||||||||||||||||||||||||||||||||||||||||||
     415 ATTAGAAGAATCTTTAACTTCCCTGTCCGAAGTGGTGCTACAGAATAGAA 366

     533 rgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeu 549
         ||||||||||||||||||||||:::   ||||||||||||||||||||||
     365 GGGGATTAGACTTATTATTCCTTAAA...GGAGGACTCTGTGCTGCCCTA 319

     550 LysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSe 566
         |||||||||||||||||||||||||||||||||||:::|||:::||||||
     318 AAAGAAGAGTGTTGTTTTTATGTTGACCATTCAGGAGTAATCAAAGATTC 269

     566 rMetAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArgGluA 583
         ||||   ||||||||:::::|||:::   |||:::||||||||||||:
     268 TATGGCCAAACTTAGAGAACGCCTAGATATACGTAAAAGAGAAAGAGAAA 219

     583 laAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMetThr 599
         :::::|||||||||||||:::||||||||:::||||||||:::|||
     218 GCCAACAAGGATGGTTCGAAAGCTGGTTTAATAAGTCCCCTTGGCTCACC 169

     600 ThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLe 616
         ||||||||||||:::::::::|||||||||:::::::|||:::||||||||
     168 ACTCTCCTCTCCACCATAGCAGGACCTTTAATTACACTTATGCTTTTGCT 119

     616 uThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluA 633
         ||||   |||||||||:::::|||:::::|||||||||:::|||||||
     118 TACTTTTGGCCCCTGCATCCTTAATAAGTTAGTAGCTTTTATTAGAGAAA 69

     633 rgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 649
         ||:::::|||||||||:::|||||||||||||||||||||::   |||
      68 GGATAAATGCAGTACAGGTTATGGTACTAAGGCAACAATATCGGGTCCTT 19

seq_name: gb_est2:BI151181

seq_documentation_block:
LOCUS       BI151181      845 bp    mRNA            EST       05-JUL-2001
DEFINITION  602916935F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5067273 5',
            mRNA sequence.
ACCESSION   BI151181
VERSION     BI151181.1  GI:14611182
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11181  row: f  column: 10
            High quality sequence stop: 774.
FEATURES              Location/Qualifiers
     source           1. .845
                      /organism="Mus musculus"
                      /strain="CZECH II (feral)"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5067273"
                      /clone_lib="NCI_CGAP_Lu29"
                      /tissue_type="spontaneous tumor, metastatic to mammary.
                      Stem cell origin."
                      /lab_host="DH10B"
                      /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigator
                      providing samples: Gilbert Smith, NIH"
BASE COUNT      225 a    188 c    216 g    215 t      1 others
ORIGIN


alignment_scores:
           Quality:  787.00            Length:    288
             Ratio:  3.422               Gaps:      8
 Percent Similarity:  79.861   Percent Identity: 58.333

alignment_block:
 US-09-171-553B-10 x BI151181   ..

 Align seg 1/1  to: BI151181  from: 1  to: 845

     309 PheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLe 325
         |||   |||:::|||   |||||||||:::|||||:::::|||||||
       2 TTTTATGCCCTTAACAGAACTGATCCAAGCGCTACTGAGGACTGCTGGCT 51

     325 uCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluArgL 342
         ||||||||||   ||||||||||||||||||||:::|||   :::
        52 ATGCCTGTCCTTGGGTCCGCCTTATTATGAAGGAATCGCCTTCAATGGAG 101

     342 ysPheAsnValThrLysGluHisArgAsnGlnCysThrTrpGlySerArg 358
           ||||||   |||::::::|||   ::::::|||:::||||||:::
     102 ATTTCAACAGAACCAGCAGCCAT...ACTTCCTGCTCTTGGGGTACAGGA 148

     359 AsnLysLeuThrLeuThrGluValSerGlyLys......GlyThrCysIl 373
         :::||||||||||||||||||||||||:::::      |||   |||||
     149 CAAAAACTGACCCTGACTGAAGTATCCGCGAGGAATCCAGGTCTCTGTAT 198

     373 eGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValT 390
         ||||      |||   :::|||:::|||||||||   :::
     199 AGGTACCCCACCTTCCACTCACAAACACCTATGCGGACAAATTCAGTCCG 248

     390 yrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrp 406
           ::::::::::::   |||   ||||||||||||:::          |||
     249 TGTCCAGAACGGAAGCTAATTACTATCTTGTACCTTCCCCGGTTGGATGG 298

     407 TrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAs 423
         ||||||||||||||||||||||||||||||||||||:::|||||||||
     299 TGGGCTTGCAATACAGGACTTACTCCCTGTGTATCAACTAAGGTTTTTAA 348

     423 nGlnSerLysAspLeuCysValMetValGlnIleValProArgValTyrT 440
         |:::|||   |||   |||||||||:::|||:::::|||||||||||||
     349 TTCATCTCATGATTT.TGTGTCATGATCCAGCTGTTACCCCGCGTATATT 397

     440 yrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArg 456
         ||||||||   :::   :::       |||   :::|||      :::|||
     398 ATCACCCTGCATCCAGTTTA......GAAGAAAGCTATGCTGGCCGGCGG 441

     457 ProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGl 473
         ||||||||||||:::::|||||||||||:::::::::|||:::||
     442 TCAAAAAGAGAACCAATTACTTTAACCCTGGCTGCATTCATGGGAATAGG 491

     473 yThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProG 490
         |   ||||||||||||||||||||||:::::|||||||||   |||   |
     492 TATGGCAGTAGGAGTGGGGACGGGAGTGTCAGCTTTGATAGAAGGAAGAC 541

     490 lnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAsp 506
         ||            |||:::   :::|||:::   |||:::::||||||
     542 AG............GGAATTCAGTCTTTGAGGGATGCTGTCAATGAAGAC 579

     507 LeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSe 523
         |||   |||:::::::::|||:::::   ||||||:::||||||||||||
     580 CTAGCGGCAATAGAGAAGTCCATTGACGCTTTAGAAAAATCTTTGACCTC 629

     523 rLeuSerGluValValLeuGlnAsnArgArgGly.LeuAspLeuLeuPhe 539
         |||||||||||||||||||||||||||||||| |||   |||||||||
     630 CCTGTCTGAGGTAGTTTTACAGAACAGGAGAGGTCTTGCATTTGTTGTTC 679

     540 LeuArgGluGlyGlyLeuCysAlaAlaLeuLysGlu.GluCysCysPheT 556
         |||:::|||||||||||||||||||||||||||||| ||||||||||||
     680 CTACAGGAAGGAGGACTGTGTGCTGCCCTTAAAGAAAGAGTGCTGCTTCT 729

     556 yrValAspHisSerGlyAlaIleArg.AspSerMetAsnLysLeuArgLy 572
         ||:::|||||:::|||   :::||| |||||||||   |||||||||::
     730 ATGCAGATCATACAGGAATAGTTAGAAGACTCTATGCCGAAACTGAGGAG 779

     572 sLys...LeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpP 588
         :|||      ::::::   :::||||||||:::|||      |||
     780 AAAGCATTCGAGCGAGGGACACCGGGAACGGGATGCTCACGGGGGTGGTG 829

     588 heGluGly 590
           :::|||
     830 AGTCCGGG 837

seq_name: gb_gss:AZ850764

seq_documentation_block:
LOCUS       AZ850764      641 bp    DNA             GSS       21-FEB-2001
DEFINITION  2M0152C18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0152C18 R, DNA sequence.
ACCESSION   AZ850764
VERSION     AZ850764.1  GI:13036088
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
```

```
           Plate: 0152  row: C  column: 18
           Seq primer: CACACAGGAAACAGCTATGACC
           Class: plasmid ends
           High quality sequence stop: 641.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0152C18"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      197 a    134 c    123 g    187 t
ORIGIN


alignment_scores:
            Quality:  770.50            Length:    218
              Ratio:  3.951               Gaps:      2
 Percent Similarity:  89.450   Percent Identity: 66.055

alignment_block:
 US-09-171-553B-10 x AZ850764/rev   ..

 Align seg 1/1  to reverse of: AZ850764  from: 1  to: 641

     417 ValSerThrSerValPheAsnGlnSerLysAspLeuCysValMetValGl 433
         ||||||   ||||||||||||:::|||||||||:::|||:::::::|||||
     639 GTGTCTATGTCTGTTTTTAATAGTTCCAAAGATTTCTGCATTTTGGTTCA 590

     433 nIleValProArgValTyrTyrHisProGluGluValValLeuAspGluT 450
         |:::::|||||:::   ||||||   ::::::      ||||||::::
     589 GCTTATTCCTAGACTCCTGTATCATGATGATAGCTCCTTTTTAGACAAAT 540

     450 yrAspTyrArgTyrAsnArgProLysArgGluProValSerLeuThrLeu 466
         ::::::::|||      |||   :::|||||||||||||:::||||||||||
     539 TTGAGCATCGGGTC...CGCTGGAGAAGAGAACCCGTTACCTTAACTTTG 493

     467 AlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGlyThrAl 483
         ||||||:::|||||||||||:::|||:::||||||||||||||||||||
     492 GCAGTTCTATTAGGATTGGGAGTAGCGGCTGGAGTAGGTACAGGAACCGC 443

     483 aAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuH 500
         ||||||||||      ||||||      |||                |||||:
     442 TGCCTTAATTAAGACCCCCCAATACTATGAA.....GAACTAC 405

     500 isAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAsn 516
         ::||||||||||      |||||||||:::::::::::|||:::::::::
     404 GTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAACCAAA 355

     517 LeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgAr 533
         ||||||||||||||||||||||||||||||:::|||||||||||||||
     354 TTAGAAGAATCTTTAACTTCCCTGTCCGAAGTGTTGCTACAGAATAGAAG 305

     533 gGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuL 550
         |||||||||||||||||||||:::|||||||||||||||||||||||||
     304 GGGATTAGACTTATTATTCCTTAAAGAAGGAGGACTCTGTGCTGCCCTAA 255

     550 ysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSer 566
         ||||||||||||||||||||||||||||||||:::|||:::||||||
     254 AAGAGGAATGTTGTTTTTATGTTGACCATTCAGGAGTAATCAAAGATTCT 205

     567 MetAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArgGluAl 583
         |||   |||||||||:::::|||:::      :::||||||||||||::
     204 ATGGCCAAACTTAGAGAACGCCTAGATATACTTAAAAGAGAAAGAGAAAG 155

     583 aAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMetThrT 600
         ::::|||||||||||||||:::|||||||||:::   ||||||:::||||
     154 CCAACAAGGATGGTTCGAAAGCTGGTTTAATAAGTTCCCTTGGCTCACCA 105

     600 hrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeu 616
         ||||||||||:::::::::|||||||||::::::|||:::|||||||||
     104 CTCTCCTCTCCACCATAGCAGGACCTTTAATTACACTTATGCTTTTGCTT 55

     617 ThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluAr 633
         |||   |||||||||:::::|||:::::|||||||||:::|||||||||
      54 ACTTTTGGCCCCTGCATCCTTAATAAGTTGGTAGCTTTTATTAGAGAAAG 5

     633 gVal 634
         |:::
       4 GATA 1

seq_name: gb_gss:AZ813042

seq_documentation_block:
LOCUS       AZ813042      625 bp    DNA              GSS       20-FEB-2001
DEFINITION  2M0080E11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0080E11 F, DNA sequence.
ACCESSION   AZ813042
VERSION     AZ813042.1  GI:12982881
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0080  row: E  column: 11
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 625.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0080E11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
```

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 191 a 128 c 120 g 186 t
ORIGIN

alignment_scores:
Quality: 757.50 Length: 207
Ratio: 4.051 Gaps: 2
Percent Similarity: 90.338 Percent Identity: 69.082

alignment_block:
US-09-171-553B-10 x AZ813042/rev ..

Align seg 1/1 to reverse of: AZ813042 from: 1 to: 625

```
408 AlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGl 424
    |||||||||:::|||||||||||||||||||||   ||||||||||||::
607 GCTTGCAATAACGGTCTCACTCCTTGTGTGTCTATGTCTGTTTTTAATAG 558

424 nSerLysAspLeuCysValMetValGlnIleValProArgValTyrTyrH 441
    :|||||||||:::|||:::::::|||||:::::|||||:::   ||||
557 TTCTAAAGATTTCTGCATTTTGGTTCAGCTTATTCCCAGACTCTTGTATC 508

441 isProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgPro 457
    ||   ::::::      ||||||:::::::::::|||      |||
507 ATGATGATAGCTCTTTTTTAGATAAATTTGAACATCGGGTC...CGCTGG 461

458 LysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyTh 474
    |||||||||||||||:::|||||||||||||||:::||||||||||||::
460 AAAAGAGAACCCGTTACCTTAACTTTGGCAGTTCTATTAGGATTGGGAGT 411

474 rAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnG 491
    :|||:::|||||||||||||||||||||||||||||      ||||||
410 AGCAGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAAGACCCCCCAAT 361

491 lnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeu 507
      |||            ||||||:::|||||||||      ||||||
360 ACTATGAA...........GAACTACGTGCAGCTATGGATATTGATCTT 323

508 ArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLe 524
    |||:::::::::::||||:::::::::|||||||||||||||||||||||
322 AGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAACTTCCCT 273

524 uSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuA 541
    |||||||||||||||||||||||||||||||||||||||||||||||:
272 GTCTGAAGTGGTGCTGCAAAATAGAAGGGGATTAGACTTATTATTCCTTA 223

541 rgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrVal 557
    ::|||||||||||||||||||||||||||||||||||||||||||||||
222 AAGAAGGAGGACTCTGTGCTGCCTTAAAAGAAGAATGTTGTTTTTATGTT 173

558 AspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLe 574
    ||||||||||||:::|||:::|||||||||   |||||||||::::::||
172 GACCATTCAGGAGTAATCAAAGATTCTATGGCTAAACTTAGAGAACGCTT 123

574 uGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyT 591
```

```
    |:::   |||:::||||||||||||:::::|||||||||||||||:::|
122 AGATATACGTAAAAGAGAAAGAGAGAGCCAACAAGGATGGTTCGAAAGCT 73

591 rpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGly 607
    ||||||||:::||||||||||:::|||||||||||||||:::::::::|||
 72 GGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACTATAGCAGGA 23

608 ProLeuValValLeuLeuLeu 614
    ||||||::::::|||:::|||
 22 CCTTTAATTACACTTATGCTT 2
```

seq_name: gb_est2:BI106669

seq_documentation_block:
LOCUS BI106669 921 bp mRNA EST 26-JUN-2001
DEFINITION 602893272F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5038336 5', mRNA sequence.
ACCESSION BI106669
VERSION BI106669.1 GI:14557562
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 921)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11105 row: p column: 17
High quality sequence stop: 805.

FEATURES Location/Qualifiers
source 1. .921
/organism="Mus musculus"
/strain="CZECH II (feral)"
/db_xref="taxon:10090"
/clone="IMAGE:5038336"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary. Stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 249 a 211 c 229 g 232 t
ORIGIN

alignment_scores:
Quality: 757.00 Length: 304
Ratio: 3.181 Gaps: 10
Percent Similarity: 78.289 Percent Identity: 53.947

alignment_block:
US-09-171-553B-10 x BI106669 ..

Align seg 1/1 to: BI106669 from: 1 to: 921

```
299 GlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSerTh 315
    ||||||:::|||:::|||:::::::|||||||||   |||::|||   ||
  4 CAGAGGATGTTTAACCTAGTGAGAGGAGCCTTTTATGCCCTTAACAGAAC 53

315 rAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProP 332
    |||||||:::||||||:::::::||||||||||||||||||||||||||||
```

```
 54 TGATCCAAGCGCTACTGAGGACTGCTGGCTATGCCTGTCCTCGGGTCCGC 103

332 roTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGlu 348
    |||||||||||||||:::|||   :::   :::||||||   |||:::::: 
104 CTTATTATGAAGGAATCGCCTTCAATGGAAATTTCAACAGAACCAGCAGC 153

349 HisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThrGl 365
    |||   ::::::|||:::|||||||:::   :::|||||||||||||||| 
154 CAT...ACTTCCTGCTCTTGGGGTACAGGACAAAAACTGACCCTGACTGA 200

365 uValSerGlyLys......GlyThrCysIleGlyLysAlaProProSerH 380
    |||||||::::::      |||   ||||||||      |||   :::| 
201 AGTATCCGCGAGGAATCCAGGTCTCTGTATAGGTACCCCACCTTCCACTC 250

380 isGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsn 396
    ||:::|||||||||   :::           ::::::::::::   ::: 
251 ACAAACACCTATGCGGACAAATTCAGTCCGTGTCCAGAACGGAAGCTGAT 300

397 GlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLe 413
       ||||||||||||:::          ||||||||||||||||||||| 
301 TACTATCTTGTACCTTCCCCGGTTGGATGGTGGGCTTGCAATACAGGACT 350

413 uThrProCysValSerThrSerValPheAsnGlnSerLysAspLeuCysV 430
    ||||||||||||||||||:::|||||||||:::|||   |||:::   | 
351 TACTCCCTGTGTATCAACTAAGGTTTTTAATTCATCTCATGATTTTCGTG 400

430 alMetValGlnIleValProArgValTyrTyrHisProGluGluValVal 446
    |||||:::|||::::::|||||||||||||||||||||||   :::   ::: 
401 TCATGATCCAGCTGTTACCCCGTGTATATTATCACCCTGCATCCAGTTTA 450

447 LeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSe 463
       |||:::|||      |||      |||   |||||||||||:::::  
451 GAAGACAGCTATGCGGGCCGG......CGGTCAAAAAGAGAACCAATTAC 494

463 rLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyT 480
    :|||||||||||:::::::::|||:::|||   |||||||||||||||| 
495 TTTAACCCTGGCTGCATTCATGGGAATAGGTATGGCAGTAGGAGTGGGGA 544

480 hrGlyThrAlaAlaLeu.Ile.ThrGlyProGlnGlnLeuGluLysGlyL 496
    |||||::::::|||::: ||| :::|||   |||           |||: 
545 CGGGAGTGTCAGCTTTCGATAGAACGGAAGACAG............GGAA 582

496 euGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGlu 512
    ::   :::|||:::   |||::::::|||||||   |||:::::::::  
583 TTCAGTCTTTGAGGGATGCTGTCAATGAAGACCTAGCGGCAATAGAGAAG 632

513 SerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLe 529
    |||::::::   ||||||:::|||||||||||||||||||||||||||| 
633 TCCATTGACGCTTTAGAAAAATCTTTGACCTCCCTGTCTGAGGTAGTTTT 682

529 uGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuC 546
    ||||||||||||||||||||||          |||:::   | 
683 ACAGAACAGGAGAGGTCTTGATTCGTCGTCCTAAAGACAGGAGGACTGTT 732

546 ysAlaAlaLeuLysGluGluCysCysPheTyr.ValAspHisSerGly.A 562
    |||||||||||||||||||||||||||||||| :::|||||||:::||| 
733 GTGCTGCCCTTAAAGAAGAGTGCTGCTTCTATTGCAGATCATACAGGAAT 782

562 laIleArgAspSer.MetAsnLysLeuArgLysLysLeuGluArgArg.. 577
      :::||||||||| |||:::::::|||   :::   ::::::   :::  
783 TAGTTAGAGACTCCTATGCAGAACCTGGACAGAACGATTTCAAGCGAAGG 832

578 .ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnA 594
        ||||||   :::|||:::::::|||      |||:::   :::| 
833 AACACGGGAACCGGATGCCCAACGGGGTAGCTGAGTCGGGTTTGAATCCA 882

594 rg 594
    ||
883 GA 884
```

seq_name: gb_gss:AZ356870

seq_documentation_block:

```
LOCUS       AZ356870      644 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0098F09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0098F09 F, DNA sequence.
ACCESSION   AZ356870
VERSION     AZ356870.1  GI:10470556
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 644)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0098  row: F  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 644.
FEATURES             Location/Qualifiers
     source          1. .644
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0098F09"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      185 a    136 c    133 g    190 t
ORIGIN
```

```
alignment_scores:
            Quality:  734.50              Length:    215
              Ratio:  3.866                 Gaps:      2
 Percent Similarity:  88.372   Percent Identity: 66.512

alignment_block:
 US-09-171-553B-10 x AZ356870/rev   ..

 Align seg 1/1  to reverse of: AZ356870  from: 1  to: 644
```

```
     391 GluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTr 407
         :::::::::::::::::::|||||||||||||||||      :::      ||
     639 CAGTCTAGTAAAAGTGGTCAGTATCTAGTGCCCCCCTCAGACACAGTATG 590

     407 pAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnG 424
         ||||||||||||||||||||||||||||||||   ||||||||||||:
     589 GGCTTGCAATACCGGTCTCACTCCTTGTGTGTCTATGTCTGTTTTTAATA 540

     424 lnSerLysAspLeuCysValMetValGlnIleValProArgValTyrTyr 440
         ::||||||:::::|||:::::||||||:::::|||||:::   |||
     539 GTTCCAAAAATTTCTGCATTTTGGTTCAGCTTATTCCTAGACTCCTGTAT 490

     441 HisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgPr 457
         |||   ::::::      ||||||:::::::::::|||      |||
     489 CATGATGATAGCTCCTTTTTAGACAAATTTGAGCATCGGGTC...CGCTG 443

     457 oLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyT 474
          :::||||||||||||:::||||||||||||||   ||||||||||||:
     442 GAGAAGAGAACCCGTTACCTTAACTTTGGCAGTTTA.TTAGGATTGGGAG 394

     474 hrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGln 490
         ::|||:::||||||||||||||||||||||||||||      ||||||
     393 TAGCGGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAAGACCCCCCAA 344

     491 GlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLe 507
             ||||||           |||:::||||||||      |||||
     343 TACTATGAAAAA.....CTACGTGCAGCTATGGATGTTGATCT 306

     507 uArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerL 524
         ||||:::::::::::|||:::::::::|||||||||||||||||||||
     305 TAGAACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAACTTCCC 256

     524 euSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeu 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     255 TGTCCGAAGTGGTGCTACAGAATAGAAGGGGATTAGACTTATTATTCCTT 206

     541 ArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrVa 557
         :::||||||||||||||||||||||||||||||||||||||||||||||||
     205 AAAGAAGGAGGACTCTGTGCTGCCCTAAAAGAAGAATGTTGTTTTTATGT 156

     557 lAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysL 574
         |||||||||||||:::|||:::|||||||||   |||||||||:::::::|
     155 TGACCATTCAGGAGTAATCAAAGATTCTATGGCCAAACTTAGAGAACGCC 106

     574 euGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGly 590
         ||:::   |||:::|||||||||||||:::::::|||||||||||||||:::
     105 TAGATATACGTAAAAGAGAAAGAGAAAGCCAACAAGGATGGTTCGAAAGC 56

     591 TrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeu 605
         |||||||||:::|||||||||:::|||||||||||||||||::::::
      55 TGGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACCATA 11

seq_name: gb_gss:AZ376517

seq_documentation_block:
LOCUS       AZ376517      610 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0130J20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0130J20 F, DNA sequence.
ACCESSION   AZ376517
VERSION     AZ376517.1  GI:10490217
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0130  row: J  column: 20
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 610.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0130J20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      177 a    131 c    120 g    182 t
ORIGIN


alignment_scores:
           Quality:  730.50            Length:    204
             Ratio:  3.970               Gaps:      2
 Percent Similarity:  90.196   Percent Identity: 67.157

alignment_block:
 US-09-171-553B-10 x AZ376517/rev   ..

 Align seg 1/1  to reverse of: AZ376517  from: 1  to: 610

     432 ValGlnIleValProArgValTyrTyrHisProGluGluValValLeuAs 448
         ||||||::::::|||||||:::   ||||||   ::::::      |||||
     608 GTTCAGCTTATTCCTAGACTCCTGTATCATGATGATAGCTCCTTTTTAGA 559

     448 pGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSerLeuT 465
         |:::::::::::|||      |||   :::||||||||||||:::||||
     558 CAGATTTGAGCATCGGGTC...CGCTGGAGAAGAGAGCCCGTTACCTTAA 512

     465 hrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGly 481
         ||||||||||:::||||||||||||:::|||:::|||||||||||||||||
     511 CTTTGGCAGTTCTATTAGGATTAGGAGTAGCGGCTGGAGTAGGTACAGGA 462

     482 ThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlyGl 498
         ||||||||||||||      ||||||      |||            ||
     461 ACCGCTGCCTTAATTAAGACCCCCCAATACTATGAA.....GA 424

     498 uLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValS 515
         ||||:::|||||||||      |||||||||:::::::::::|||::::
```

```
     423 ACTACGTGCAGCTATGGATGTTGATCTTAGAACTATAGAACAGTCTATAA 374

     515 erAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsn 531
         :::::|||||||||||||||||||||||||||||||||||||||||||||
     373 CCAAATTAGAAGAATCTTTAACTTCCCTGTCCGAAGTGGTGCTACAGAAT 324

     532 ArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAl 548
         ||||||||||||||||||||||||||:::||||||||||||||||||||
     323 AGAAGGGGATTAGACTTATTATTCCTTAAAGAAGGAGGACTCTGTGCTGC 274

     548 aLeuLysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArgA 565
         |||||||||||||||||||||||||||||||||||||:::|||:::|
     273 CCTAAAAGAAGAATGTTGTTTTTATGTTGACCATTCAGGAGTAATCAAAG 224

     565 spSerMetAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArg 581
         ||||||||   |||||||||::::::|||:::   |||:::|||||||||
     223 ATTCTATGGCCAAACTTAGAGAACGCCTAGATATACGTAAAAGAGAAAGA 174

     582 GluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMe 598
         |||::::::|||||||||||||||:::|||||||||:::||||||||::
     173 GAAAGCCAACAAGGATGGTTCGAAAGCTGGTTTAATAAGTCCCCTTGGCT 124

     598 tThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuL 615
         :|||||||||||||||:::::::::|||||||||::::::|||:::||||
     123 CACCACTCTCCTCTCCACCATAGCAGGACCTTTGATTACACTTATGCTTT 74

     615 euLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArg 631
         ||||||||   |||||||||::::::|||::::::|||||||||:::|||
      73 TGCTTACTTTTGGCCCCTGCATCCTTAATAAGTTAGTAGCTTTTATTAGA 24

     632 GluArgValSer 635
         ||||||::::::
      23 GAAAGGATAAAT 12

seq_name: gb_gss:AZ430249

seq_documentation_block:
LOCUS       AZ430249      591 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0214G11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0214G11 R, DNA sequence.
ACCESSION   AZ430249
VERSION     AZ430249.1  GI:10554262
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0214  row: G  column: 11
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 591.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0214G11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      165 a    125 c    115 g    186 t
ORIGIN


alignment_scores:
            Quality:  729.00             Length:    194
              Ratio:  4.096                Gaps:      1
 Percent Similarity:  91.753   Percent Identity: 71.134

alignment_block:
 US-09-171-553B-10 x AZ430249/rev   ..

 Align seg 1/1  to reverse of: AZ430249  from: 1  to: 591

     456 ArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLe 472
         |||   |||||||||||||||:::|||||||||||||||::::|||||||
     588 CGCTGGAAAAGAGAACCCGTTACCTTAACTTTGGCAGTTTTATTAGGATT 539

     472 uGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyP 489
         ||||:::|||:::|||||||||||||||||||||||||||||||:::   |
     538 GGGAGTAGCAGCTGGAGTAGGTACAGGAACCGCTGCCTTAATTAACACCC 489

     489 roGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGlu 505
         |||||      |||            ||||||:::|||||||||
     488 CCCAATACTATGAA............GAGCTACGTGCAGCTATGGATGTT 451

     506 AspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuTh 522
         |||||||||:::::::::::::|||:::::::::|||||||||||||||||
     450 GATCTTAGAACCATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAAC 401

     522 rSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuP 539
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     400 TTCCCTATCCGAAGTGGTGCTGCAAAATAGAAGGGGGTTAGACTTATTAT 351

     539 heLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPhe 555
         |||||:::|||||||||||||||||    |||||||||||||||||||||
     350 TCCTTAAAGAAGGAGGACTCTGTGCTCCCTTAAAAGAAGAATGTTGTTTT 301

     556 TyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLy 572
         ||||||||||||||||||:::|||:::|||||||||   |||||:::::
     300 TATGTTGACCATTCAGGAGTAATCAAAGATTCTATGGCTAAACTTAAAGA 251

     572 sLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheG 589
         :::|||:::   |||:::|||||||||||||:::   ||||||||||||
     250 ACGCCTAGATATACGTAAAAGAGAAAGAGAAAGCCCACAAGGATGGTTTG 201

     589 luGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeu 605
         ||:::|||||||||:::|||||||||:::|||||||||||||||:::::::
     200 AAAGCTGGTTTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACTATA 151
```

```
     606 ThrGlyProLeuValValLeuLeuLeuLeuThrValGlyProCysLe 622
         :::|||||||||:::::|||:::|||||||||||   ||||||||::
     150 GCAGGACCTTTAATTACACTTATGCTTTTGCTTACTTTTGGCCCATGCAT 101

     622 uIleAsnArgPheValAlaPheValArgGluArgValSerAlaValGlnI 639
         ::::|||:::::::|||||||||:::|||||||||:::::|||||||||:
     100 CCTTAATAAGTTAGTAGCTTTTATTAGAGAAAGGATAAACGCAGTCCAAG 51

     639 leMetValLeuArgGlnGlnTyrGlnGlyLeu 649
         ::|||||||||||||||||||:::   |||
      50 TTATGGTACTAAGGCAACAATATCGGGTCCTT 19

seq_name: gb_gss:AZ642336

seq_documentation_block:
LOCUS       AZ642336      624 bp    DNA             GSS       14-DEC-2000
DEFINITION  1M0505M18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0505M18 F, DNA sequence.
ACCESSION   AZ642336
VERSION     AZ642336.1  GI:11768843
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0505  row: M  column: 18
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 624.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0505M18"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      203 a    116 c    131 g    174 t
ORIGIN


alignment_scores:
           Quality:  721.50            Length:    197
             Ratio:  3.964               Gaps:      2
 Percent Similarity:  92.386   Percent Identity: 70.051

alignment_block:
 US-09-171-553B-10 x AZ642336   ..

 Align seg 1/1  to: AZ642336  from: 1  to: 624

     461 ProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGl 477
         ||||||:::||||||||||||||||||:::|||||||||||:::|||:::||
       3 CCCGTTACCTTAACTTTGGCAGTTCTATTAGGATTGGGAGTAGCAGCTGG 52

     477 yValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluL 494
         ||||||||||||||||||||||||||||       ||||||      |||
      53 AGTAGGTACAGGAACCGCTGCCTTAATTAAGACCCCCCAATACTATGAA. 101

     494 ysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeu 510
                   ||||||:::|||||||||       |||||||||::::::
     102 ..........GAACTACGTGCAGCTATGGATATTGATCTTAGAACTATA 140

     511 LysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluVa 527
         ::::::|||:::::::::|||||||||||||||||||||||||||||||||
     141 GAACAGTCTATAACCAAATTAGAAGAATCTTTAACTTCCCTGTCCGAAGT 190

     527 lValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyG 544
         |||||||||||||||||||||||||||||||||||||:::|||||||
     191 GGTGCTGCAAAATAGAAGGGGATTAGACTTATTATTCCTTAAAGAAGGAG 240

     544 lyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAspHisSer 560
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     241 GACTCTGTGCTGCCTTAAAAGAAGAATGTTGTTTTTATGTTGACCATTCA 290

     561 GlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGluArgAr 577
         |||:::|||:::|||||||||   |||||||||::::::|||:::   ||
     291 GGAGTAATCAAAGATTCTATGGCTAAACTTAGAGAACGCCTAGATATACG 340

     577 gArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnA 594
         |:::|||||||||||||:::::::|||||||||||||||:::|||||||||:
     341 TAAAAGAGAAAGAGAAAGCCAACAAGGATGGTTTGAAAGCTGGTTTAATA 390

     594 rgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuVal 610
         ::|||||||||:::|||||||||||||||:::::::::|||||||||:::
     391 AGTCCCCTTGGCTCACCACTCTCCTCTCCACTATAGCAGGACCTTTAATT 440

     611 ValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheVa 627
         :::|||:::|||||||||||||   |||||||||::::::|||:::::||
     441 ACACTTATGCTTTTGCTTACTTTTGGCCCATGCATCCTTAATAAGTTAGT 490

     627 lAlaPheValArgGluArgValSerAlaValGlnIleMetValLeuArgG 644
         |||||||:::|||||||||:::::::|||||||||:::|||||||||||||
     491 AGCTTTTATTAGAGAAAGGATAAACGCAGTCCAAGTTATGGTCCTAAGAC 540

     644 lnGlnTyrGlnGlyLeuLeuSerGlnGlyGluThrAspLeu 657
         ||||||||:::   :::|||::::::   |||:::::|||
     541 AACAATATCGG...GTCCTTCAGGAGGTTGAAAACTCGCTC 578

seq_name: gb_gss:AZ337993

seq_documentation_block:
LOCUS       AZ337993      673 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0068F24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0068F24 R, DNA sequence.
ACCESSION   AZ337993
VERSION     AZ337993.1  GI:10410826
KEYWORDS    GSS.
SOURCE      house mouse.
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0068  row: F  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 673.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0068F24"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      197 a    120 c    167 g    188 t      1 others
ORIGIN


alignment_scores:
            Quality:  721.50            Length:    228
              Ratio:  3.607               Gaps:      3
 Percent Similarity:  87.719   Percent Identity: 58.772

alignment_block:
 US-09-171-553B-10 x AZ337993   ..

 Align seg 1/1  to: AZ337993  from: 1  to: 673

     421 ValPheAsnGlnSerLysAspLeuCysValMetValGlnIleValProAr 437
         |||||||||:::|||   |||:::||||||||||:::|||:::::::|||||
       1 GTTTTTAATTCATCTCATGATTTTTGTGTCATGATCCAGCTGTTACCCCG 50

     437 gValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyrArgT 454
         |||||||||||||||   :::   :::      |||   :::|||
      51 CGTATATTATCACCCTGCATCCAGTTTA......GAAGAAAGCTATGCTG 94

     454 yrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeu 470
         :::|||   ||||||||||||:::::|||||||||||||:::::::::
      95 GCCGGCGGTCAAAAAGAGAACCAATTACTTTAACCCTGGCTGCATTCATG 144

     471 GlyLeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleTh 487
         |||:::|||   |||||||||||||||||||:::::|||||||||
     145 GGAATAGGTATGGCAGTAGGAGTGGGGACGGGAGTGTCAGCTTTGATAGA 194

     487 rGlyProGlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetT 504
          |||   |||            |||:::   :::|||:::   |||::::
     195 AGGAAGACAG.............GGAATTCAGTCTTTGAGGGATGCTGTCA 232

     504 hrGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSer 520
         ::|||||||||   |||:::::::::|||:::::::   ||||||:::|||
     233 ATGAAGACCTAGCAGCAATAGAGAAGTCCATTGACGCTTTAGAAAAATCT 282

     521 LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLe 537
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     283 TTGACCTCCCTGTCTGAGGTAGTTTTACAGAACAGGAGAGGTCTTGATTT 332

     537 uLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysC 554
         |||||||||:::||||||||||||||||||||||||||||||||||||||
     333 GTTGTTCCTAAAGGAAGGAGGACTGTGTGCTGCCCTTAAAGAAGAGTGCT 382

     554 ysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeu 570
         ||||||||:::|||||:::|||   :::|||||||||||||:::||||||
     383 GCTTCTATGCAGATCATACAGGAATAGTTAGAGACTCTATGCAGAAACTG 432

     571 ArgLysLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTr 587
         |||:::::::||||||:::|||:::|||||||||:::|||:::::::||||
     433 AGAGAAAGATTAGAGCAAAGGAAACGGGAACGGGATGCTCAACGGGGGTG 482

     587 pPheGluGlyTrpPheAsnArgSerPro...TrpMetThrThrLeuLeuS 603
         |||||||:::|||||||:::   :::|||   |||:::|||:::|||:::|
     483 GTTTGAGTCGTGGTTTGAATCACAACCATCTTGGATAACTTCTTTAATTT 532

     603 erAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThrValGly 619
         ||||:::::::||||||:::::::::::::   |||       :::   |||
     533 CCGCTGTAGCCGGACCAATCCTTATGATATGCTTAGCTNTAGTTTTCGGC 582

     620 ProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAl 636
         ||||||:::|||||||||   :::||||||:::::::::::::::::::::
     583 CCTTGTATAATAAATAGAGGAATGGCTTTCATCCAGAGTAAAATTGATAC 632

     636 aValGlnIleMetValLeuArgGlnGlnTyrGln 647
         :|||:::::::|||||||||:::::::|||||||||
     633 AGTAAAACTCATGGTTCTTCAAAGGCAATATCAA 666

seq_name: gb_gss:AZ814037

seq_documentation_block:
LOCUS       AZ814037      632 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0081N06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0081N06 R, DNA sequence.
ACCESSION   AZ814037
VERSION     AZ814037.1  GI:12983945
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
```

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0081 row: N column: 06
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 632.

FEATURES Location/Qualifiers
source 1. .632
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0081N06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 203 a 115 c 138 g 176 t
ORIGIN

alignment_scores:
Quality: 713.50 Length: 199
Ratio: 3.920 Gaps: 2
Percent Similarity: 91.457 Percent Identity: 68.844

alignment_block:
US-09-171-553B-10 x AZ814037 ..

Align seg 1/1 to: AZ814037 from: 1 to: 632

```
459 ArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAl 475
    ||||||||||||:::|||||||||||||||:::|||||||||||||:::::
  3 AGAGAACCCGTTACCTTAACTTTGGCAGTTCTATTAGGATTGGGAGTAGT 52

475 aValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnL 492
    ::::||||||||||||||||||||||||||||      ||||||
 53 GGCTGGAGTAGGTACAGGAACTGCTGCCTTAATTAAGACCCCCCAATACT 102

492 euGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArg 508
      |||          ||||||:::|||||||||      |||||||||
103 ATGAA.....GAACTACGTGCAGCTATGGATGTTGATCTTAGA 140

509 AlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSe 525
    ::::::::::::|||:::::::::|||||||||||||||||||||||||
141 ACTATAGAACAGTCTATAACCAAATTAGAAGAATCTTTAACTTCCCTGTC 190

525 rGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgG 542
    |||||||||||||||||||||||||||||||||||||||||||||:::|
191 CGAAGTGGTGCTACAGAATAGAAGGGGATTAGACTTATTATTCCTTAAAG 240

542 luGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAsp 558
    ||||||||||||||||||||||||||||   ||||||||||||||||||
241 AAGGAGGACTCTGTGCTGCCCTAAAAGAAGGATGTTGTTTTTATGTTGAC 290

559 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGl 575
    |||||||||:::|||:::|||||||||   |||||||||:::::|||::
291 CATTCAGGAGTAATCAAAGATTCTATGGCCAAACTTAGAGAACGCCTAGA 340

575 uArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpP 592
    :   |||:::|||||||||||||:::::|||||||||||||||||:::||||
341 TATACGTAAAAGAGAAAGAGAAAGCCAACAAGGATGGTTCGAAAGCTGGT 390

592 heAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyPro 608
    |||||:::|||||||||:::|||||||||||||||:::::::::||||||
391 TTAATAAGTCCCCTTGGCTCACCACTCTCCTCTCCACCATAGCAGGACCT 440

609 LeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnAr 625
    |||:::::::|||:::||||||||||||   |||   |||::::::|||:
441 TTAATTACACTTATGCTTTTGCTTACTTTTGGCTCCTGCATCCTTAATAA 490

625 gPheValAlaPheValArgGluArgValSerAlaValGlnIleMetValL 642
    ::::||||||||:::||||||||:::::::|||||||:::|||||||
491 GTTAGTCGCTTTTATTAGAGAAAGGATAAATGCAGTACAGGTTATGGTAC 540

642 euArgGlnGlnTyrGlnGlyLeuLeuSerGlnGlyGluThrAspLeu 657
    |||||||||||||::   :::|||::::::  |||:::::|||
541 TAAGGCAACAATATCGG...GTCCTTCAAGAGGTTGAAAACTCGCTC 584
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:37:18 ; Search time 6550.43 Seconds
(without alignments)
32.809 Million cell updates/sec

Title: US-09-171-553B-13
Perfect score: 20
Sequence: 1 gatggctctcctgccctttg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18.4	92.0	391	11	BF465634	BF465634 UI-M-CG0p
	2	18.4	92.0	429	10	AA797871	AA797871 vw31h06.r
	3	18.4	92.0	435	11	BF565386	BF565386 UI-R-BO1-
	4	18.4	92.0	444	10	AA472060	vg97d05.r
	5	18.4	92.0	488	11	BF719561	mab41e06.
	6	18.4	92.0	510	10	AA033199	AA033199 mi36d04.r
c	7	18.4	92.0	539	13	AQ512916	HS_5139_A
c	8	18.4	92.0	566	10	AA121599	AA121599 zn78a06.r
c	9	18.4	92.0	571	10	BE097309	BE097309 UI-R-BO1-
c	10	18.4	92.0	592	10	AW556515	AW556515 L0269D06-
c	11	18.4	92.0	600	10	AW488316	AW488316 UI-M-BH3-
	12	18.4	92.0	659	10	BE374286	BE374286 601227645
c	13	18.4	92.0	690	11	BG075579	BG075579 H3149A02-
	14	18.4	92.0	896	11	BG342613	BG342613 602374796
	15	18.4	92.0	2171	12	AK019522	AK019522 Mus muscu
	16	17.4	87.0	257	10	AW326534	AW326534 19216 MAR
c	17	17.4	87.0	397	10	AW489995	AW489995 UI-M-BH3-
	18	17.4	87.0	455	10	AI013028	AI013028 EST207479
c	19	17.4	87.0	581	13	AZ777840	AZ777840 2M0012D10
c	20	17.4	87.0	607	11	W20084	W20084 zb40f01.r1
c	21	17.4	87.0	783	13	BH051324	BH051324 RPCI-24-3
	22	17	85.0	691	10	AI735229	AI735229 at08a01.x
c	23	17	85.0	1071	10	BE612516	BE612516 601452020
c	24	17	85.0	1787	11	BG256691	BG256691 602370880
	25	16.8	84.0	190	11	T25876	T25876 ESTDIR3 CD3
c	26	16.8	84.0	219	10	BB362510	BB362510 BB362510
c	27	16.8	84.0	250	10	AA478651	AA478651 zv19c05.r
c	28	16.8	84.0	268	10	AV228430	AV228430 AV228430
	29	16.8	84.0	422	13	AQ902607	AQ902607 LMAJFV1_1
c	30	16.8	84.0	454	11	BF549019	BF549019 UI-R-A0-a
	31	16.8	84.0	473	10	AA170346	AA170346 ms88g09.r
c	32	16.8	84.0	492	11	BF554951	BF554951 UI-R-E0-c
	33	16.8	84.0	504	13	AZ582584	AZ582584 1M0376O03
	34	16.8	84.0	536	13	AQ456957	AQ456957 HS_5151_A
c	35	16.8	84.0	539	10	BE586605	BE586605 WHE0509_G
	36	16.8	84.0	555	10	BE650111	BE650111 UI-M-BH3-
c	37	16.8	84.0	558	13	AQ523881	AQ523881 HS_5246_A
	38	16.8	84.0	576	13	AZ730616	AZ730616 RPCI-24-1
c	39	16.8	84.0	594	13	AZ951110	AZ951110 2M0215M20
	40	16.8	84.0	602	10	AI102103	AI102103 EST211392
	41	16.8	84.0	610	13	AZ432720	AZ432720 1M0218E17
	42	16.8	84.0	610	13	AZ859488	AZ859488 2M0165D07
c	43	16.8	84.0	649	13	AZ741186	AZ741186 RPCI-24-6
	44	16.8	84.0	654	13	AZ453887	AZ453887 1M0255N03
c	45	16.8	84.0	658	13	AZ457973	AZ457973 1M0261A07

ALIGNMENTS

RESULT 1
BF465634
LOCUS BF465634 391 bp mRNA EST 04-DEC-2000
DEFINITION UI-M-CG0p-bql-h-12-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone UI-M-CG0p-bql-h-12-0-UI 3', mRNA sequence.
ACCESSION BF465634
VERSION BF465634.1 GI:11534817
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 391)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=No.
FEATURES Location/Qualifiers

```
     source          1. .391
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-CG0p-bql-h-12-0-UI"
                     /clone_lib="NIH_BMAP_Ret4_S2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_Ret4_S2 library is a subtracted library,
                     ultimately derived from mouse retina tissue libraries at
                     various stages of development. For a detailed description
                     of the library from which this clone was derived,  please
                     visit our web site at brainest.eng.uiowa.edu.
                     TAG_SEQ=None found"
BASE COUNT       88 a    100 c    123 g     80 t
ORIGIN


  Query Match             92.0%;  Score 18.4;  DB 11;  Length 391;
  Best Local Similarity   95.0%;  Pred. No. 4.4e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccctttg 20
          |||||||||||| |||||||
Db    139 GATGGCTCTCCTTCCCTTTG 158


RESULT   2
AA797871
LOCUS       AA797871      429 bp    mRNA            EST       10-FEB-1998
DEFINITION  vw31h06.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1245467 5', mRNA sequence.
ACCESSION   AA797871
VERSION     AA797871.1  GI:2860826
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:659155
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 358.
FEATURES             Location/Qualifiers
     source          1. .429
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1245467"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of  the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT       95 a    109 c    128 g     97 t
ORIGIN


  Query Match             92.0%;  Score 18.4;  DB 10;  Length 429;
  Best Local Similarity   95.0%;  Pred. No. 4.4e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccctttg 20
          |||||||||||| |||||||
Db     13 GATGGCTCTCCTTCCCTTTG 32


RESULT   3
BF565386
LOCUS       BF565386      435 bp    mRNA            EST       12-DEC-2000
DEFINITION  UI-R-BO1-ajv-e-06-0-UI.r1 UI-R-BO1 Rattus norvegicus cDNA clone
            UI-R-BO1-ajv-e-06-0-UI 5', mRNA sequence.
ACCESSION   BF565386
VERSION     BF565386.1  GI:11675116
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1798920
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BO1-ajv-e-06-0-UI"
                     /clone_lib="UI-R-BO1"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The library
                     (UI-R-BO1) is a subtracted library derived from a mixture
                     of the following tissues: thalamus, cerebellum,
                     hypothalamus, medulla, pons, midbrain, cerebral cortex,
                     corpus striatum and hippocampus. For a detailed
                     description of the library from which this clone was
                     derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)"
```

```
BASE COUNT      105 a    121 c    115 g     94 t
ORIGIN


  Query Match             92.0%;  Score 18.4;  DB 11;  Length 435;
  Best Local Similarity   95.0%;  Pred. No. 4.4e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccctttg 20
          |||||||||||| |||||||
Db    346 GATGGCTCTCCTTCCCTTTG 365


RESULT   4
AA472060
LOCUS       AA472060      444 bp    mRNA            EST       18-JUN-1997
DEFINITION  vg97d05.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:873897 5' similar to TR:G190352 G190352 PEREGRIN. ;, mRNA
            sequence.
ACCESSION   AA472060
VERSION     AA472060.1  GI:2200051
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 444)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:513377
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .444
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:873897"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      104 a    116 c    127 g     97 t
ORIGIN


  Query Match             92.0%;  Score 18.4;  DB 10;  Length 444;
  Best Local Similarity   95.0%;  Pred. No. 4.4e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccctttg 20
          |||||||||||| |||||||
Db    301 GATGGCTCTCCTTCCCTTTG 320


RESULT   5
BF719561
LOCUS       BF719561      488 bp    mRNA            EST       03-JAN-2001
DEFINITION  mab41e06.y1 Soares_NMEBA_branchial_arch Mus musculus cDNA clone
            IMAGE:3972947 5' similar to SW:BR14_HUMAN P55201 PEREGRIN ;, mRNA
            sequence.
ACCESSION   BF719561
VERSION     BF719561.1  GI:12020563
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1472979
            Seq primer: -40RP from Gibco
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3972947"
                     /clone_lib="Soares_NMEBA_branchial_arch"
                     /tissue_type="branchial arches"
                     /dev_stage="embryo, 10.5 dpc"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATGCATTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      115 a    135 c    128 g    110 t
ORIGIN


  Query Match             92.0%;  Score 18.4;  DB 11;  Length 488;
  Best Local Similarity   95.0%;  Pred. No. 4.5e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccctttg 20
          |||||||||||| |||||||
Db    425 GATGGCTCTCCTTCCCTTTG 444


RESULT   6
AA033199
LOCUS       AA033199      510 bp    mRNA            EST       22-AUG-1996
DEFINITION  mi36d04.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
```

```
            clone IMAGE:465607 5' similar to PIR:JC2069 JC2069 zinc-finger
            protein, BR140 - human ;, mRNA sequence.
ACCESSION   AA033199
VERSION     AA033199.1  GI:1504601
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:279423
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 505.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:465607"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT      121 a    135 c    142 g    112 t
ORIGIN


  Query Match             92.0%;  Score 18.4;  DB 10;  Length 510;
  Best Local Similarity   95.0%;  Pred. No. 4.5e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gatggctctcctgccctttg 20
           |||||||||||| |||||||
Db     408 GATGGCTCTCCTTCCCTTTG 427


RESULT   7
AQ512916/c
LOCUS       AQ512916      539 bp    DNA             GSS       05-MAY-1999
DEFINITION  HS_5139_A2_B09_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=715 Col=18 Row=C, DNA sequence.
ACCESSION   AQ512916
VERSION     AQ512916.1  GI:4745207
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 715  row: C  column: 18
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 539.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=715 Col=18 Row=C"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      147 a    179 c     99 g    108 t      6 others
ORIGIN


  Query Match             92.0%;  Score 18.4;  DB 13;  Length 539;
  Best Local Similarity   95.0%;  Pred. No. 4.5e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gatggctctcctgccctttg 20
           | ||||||||||||||||||
Db     135 GTTGGCTCTCCTGCCCTTTG 116


RESULT   8
AA121599/c
LOCUS       AA121599      566 bp    mRNA            EST       19-NOV-1996
DEFINITION  zn78a06.r1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
            cDNA clone IMAGE:564274 5', mRNA sequence.
ACCESSION   AA121599
VERSION     AA121599.1  GI:1679214
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 566)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
```

```
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 439.
FEATURES             Location/Qualifiers
     source          1. .566
                     /organism="Homo sapiens"
                     /db_xref="GDB:4596797"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:564274"
                     /clone_lib="Stratagene NT2 neuronal precursor 937230"
                     /tissue_type="neuroepithelial cells"
                     /dev_stage="Ntera-2 neuroepithelial cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: brain; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Uninduced, exponentially growing neuroepithelial
                     cells (Ntera-2/c1.D1). Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      114 a    157 c    141 g    151 t      3 others
ORIGIN
```

```
  Query Match             92.0%;  Score 18.4;  DB 10;  Length 566;
  Best Local Similarity   95.0%;  Pred. No. 4.5e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccctttg 20
          || ||||||||||||||||||
Db    348 GAGGGCTCTCCTGCCCTTTG 329
```

```
RESULT   9
BE097309/c
LOCUS       BE097309      571 bp    mRNA            EST       12-JUN-2000
DEFINITION  UI-R-BO1-apv-g-12-0-UI.s1 UI-R-BO1 Rattus norvegicus cDNA clone
            UI-R-BO1-apv-g-12-0-UI 3', mRNA sequence.
ACCESSION   BE097309
VERSION     BE097309.1  GI:8488240
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized medulla library cDNA Library Preparation: M.B. Soares
            Lab Clone distribution: clones will be available through Research
            Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .571
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BO1-apv-g-12-0-UI"
                     /clone_lib="UI-R-BO1"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The library
                     (UI-R-BO1) is a subtracted library derived from a mixture
                     of the following tissues: thalamus, cerebellum,
                     hypothalamus, medulla, pons, midbrain, cerebral cortex,
                     corpus striatum and hippocampus. For a detailed
                     description of the library from which this clone was
                     derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_LIB=UI-R-BO1
                     TAG_TISSUE=medulla
                     TAG_SEQ=GAACCG"
BASE COUNT      143 a    152 c    134 g    141 t      1 others
ORIGIN
```

```
  Query Match             92.0%;  Score 18.4;  DB 10;  Length 571;
  Best Local Similarity   95.0%;  Pred. No. 4.5e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccctttg 20
          |||||||||||| |||||||
Db    528 GATGGCTCTCCTTCCCTTTG 509
```

```
RESULT  10
AW556515/c
LOCUS       AW556515      592 bp    mRNA            EST       31-AUG-2000
DEFINITION  L0269D06-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA
            clone L0269D06 3', mRNA sequence.
ACCESSION   AW556515
VERSION     AW556515.1  GI:7201944
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
            ,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
            III, Becker,K.G. and Ko,M.S.H.
  TITLE     Genome-wide expression profiling of mid-gestation placenta and
            embryo using a 15,000 mouse developmental cDNA microarray
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
  MEDLINE   20381348
COMMENT     Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: L0269  row: D  column: 06
            Seq primer: -21M13 Forward
            High quality sequence stop: 592
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Mus musculus"
```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L0269D06"
/clone_lib="NIA Mouse Newborn Ovary cDNA Library"
/sex="female"
/dev_stage="Newborn Ovary"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Gibco/BRL Life Technology); Site_1: SalI; Site_2: NotI; Total RNAs were extracted from 7 Newborn Ovary. The double-stranded cDNA was synthesized by Gibco's kit with an Oligo(dT) primer [NotI primer-adapter from GibcoBRL] [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from 2.56ug of total RNA . The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal3 (include Sal1 sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang and Yulan Piao."
BASE COUNT 140 a 166 c 146 g 137 t 3 others
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 592;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gatggctctcctgccctttg 20
|||||||||||| |||||||
Db 511 GATGGCTCTCCTTCCCTTTG 492

RESULT 11
AW488316/c
LOCUS AW488316 600 bp mRNA EST 24-FEB-2000
DEFINITION UI-M-BH3-arp-b-06-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone UI-M-BH3-arp-b-06-0-UI 3', mRNA sequence.
ACCESSION AW488316
VERSION AW488316.1 GI:7058586
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 600)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized cerebellum library cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .600
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-arp-b-06-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=cerebellum
TAG_SEQ=GACTC"
BASE COUNT 141 a 166 c 143 g 150 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 600;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gatggctctcctgccctttg 20
|||||||||||| |||||||
Db 528 GATGGCTCTCCTTCCCTTTG 509

RESULT 12
BE374286
LOCUS BE374286 659 bp mRNA EST 21-JUL-2000
DEFINITION 601227645F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3585899 5', mRNA sequence.
ACCESSION BE374286
VERSION BE374286.1 GI:9319649
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 659)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8746  row: j  column: 12
            High quality sequence stop: 586.
FEATURES              Location/Qualifiers
     source           1. .659
                      /organism="Mus musculus"
                      /strain="FVB/N"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3585899"
                      /clone_lib="NCI_CGAP_Mam1"
                      /tissue_type="tumor, biopsy sample"
                      /dev_stage="3 months, virgin"
                      /lab_host="DH10B"
                      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigator
                      providing samples: Gilbert Smith, NIH"
BASE COUNT      149 a    178 c    189 g    143 t
ORIGIN


  Query Match             92.0%;  Score 18.4;  DB 10;  Length 659;
  Best Local Similarity   95.0%;  Pred. No. 4.5e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccctttg 20
          |||||||||||| |||||||
Db    452 GATGGCTCTCCTTCCCTTTG 471


RESULT  13
BG075579/c
LOCUS       BG075579      690 bp    mRNA            EST       26-JAN-2001
DEFINITION  H3149A02-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3149A02 3', mRNA sequence.
ACCESSION   BG075579
VERSION     BG075579.1  GI:12558148
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3149A02-5
            Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3149  row: A  column: 02
            Seq primer: -21M13 Forward
            High quality sequence stop: 690
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .690
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="H3149A02"
                      /clone_lib="NIA Mouse 15K cDNA Clone Set"
                      /sex="Clones arrayed from a variety of cDNA libraries"
                      /dev_stage="Clones arrayed from a variety of cDNA
                      libraries"
                      /lab_host="DH10B"
                      /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                      clone is among a rearrayed set of 15,247 clones from 11
                      embryo  cDNA libraries (including preimplantation stage
                      embryos from  unfertilized egg to blastocyst, embryonic
                      part of E7.5 embryos,  extraembryonic part of E7.5 embryos
                      , and E12.5 female mesonephros/gonad)  and one newborn
                      ovary cDNA library. Average insert size 1.5 kb. All
                      source libraries are cloned unidirectionally with Oligo(dT
                      )-Not primers.  References include: (1) Genome-wide
                      expression profiling of mid-gestation  placenta and embryo
                      using a 15,000 mouse developmental cDNA microarray,  2000,
                      Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                      Large-scale cDNA  analysis reveals phased gene expression
                      patterns during preimplantation  mouse develolpment, 2000,
                      Development, 127: 1737-1749; (3) Genome-wide  mapping of
                      unselected transcripts from extraembryonic tissue of
                      7.5-day  mouse embryos reveals enrichment in the t-complex
                      and under-representation  on the X chromosome, 1998, Hum
                      Mol Genet 7: 1967-1978."
BASE COUNT      158 a    197 c    170 g    165 t
ORIGIN


  Query Match             92.0%;  Score 18.4;  DB 11;  Length 690;
  Best Local Similarity   95.0%;  Pred. No. 4.5e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gatggctctcctgccctttg 20
          |||||||||||| |||||||
Db    512 GATGGCTCTCCTTCCCTTTG 493


RESULT  14
BG342613
LOCUS       BG342613      896 bp    mRNA            EST       27-FEB-2001
DEFINITION  602374796F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4482310 5',
            mRNA sequence.
ACCESSION   BG342613
VERSION     BG342613.1  GI:13149051
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 896)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10318  row: p  column: 23
            High quality sequence stop: 564.
FEATURES              Location/Qualifiers
     source           1. .896
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:4482310"
                      /clone_lib="NIH_MGC_94"
                      /tissue_type="retina"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
```

Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 243 a 259 c 228 g 166 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 11; Length 896;
Best Local Similarity 95.0%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gatggctctcctgccctttg 20
||||||||||||| |||||||
Db 138 GATGGCTCTCCTTCCCTTTG 157

RESULT 15
AK019522
LOCUS AK019522 2171 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched library, clone:4833438B11, full insert sequence.
ACCESSION AK019522
VERSION AK019522.1 GI:12859783
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 0 day neonate head cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:4833438B11.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2171)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 2171)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 2171)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome research. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (bases 1 to 2171)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 2171)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 0.5 and 3 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
FEATURES Location/Qualifiers
source 1. .2171
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1912802"
/db_xref="MGD:MGI:1926033"
/clone="4833438B11"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
CDS <1. .1477
/note="putative"
/codon_start=2
/protein_id="BAB31777.1"
/db_xref="GI:12859784"
/translation="EQGGAVLRQARRQAEKMGIDFETGMHIPHNLAGDEVSHHTEDVE
EERLVLLENQKHLPVEEQLKLLLERLDEVNASKQSVGRSRRAKMIKKEMTALRRKLAH
QRETGRDGPERHGPSGRGNLTPHPAACDKDGQTDSAAEESSSQETSKGLGPNMSSTPA
HEVGRRTSVLFSKKNPKTAGPPKRPGRPPKNRESQMTPSHGGSPVGPPQLPIMGSLRQ
RKRGRSPRPSSSSDSDSDKSTEDPPMDLPANGFSSGNQPVKKSFLVYRNDCNLPRSSS
DSESSSSSSSSAASDRTSTTPSKQGRGKPSFSRGTFPEDSSEDTSGTENEAYSVGTGR
GVGHSMVRKSLGRGAGWLSEDEDSPLDALDLVWAKCRGYPSYPALIIDPKMPREGMFH
HGVPIPVPPLEVLKLGEQMTQEAREHLYLVLFFDNKRTWQWLPRTKLVPLGVNQDLDK
EKMLEGRKSNIRKSVQIAYHRALQHRSKVQGEQSSETSDSD"
polyA_signal 2151. .2156
/note="putative"
polyA_site 2171
/note="putative"
BASE COUNT 573 a 577 c 598 g 423 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 12; Length 2171;
Best Local Similarity 95.0%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gatggctctcctgccctttg 20
||||||||||| |||||||
Db 1503 GATGGCTCTCCTTCCCTTTG 1522

Search completed: February 23, 2002, 23:37:21
Job time: 9442 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:37:21 ; Search time 6550.43 Seconds
(without alignments)
29.528 Million cell updates/sec

Title: US-09-171-553B-14
Perfect score: 18
Sequence: 1 ccacagtcgtacaccacg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	18	100.0	558	11	BI360879	BI360879 388957 MA
	2	17	94.4	323	13	AZ056887	AZ056887 RPCI-23-4
c	3	17	94.4	559	13	AZ749422	AZ749422 RPCI-24-1
	4	17	94.4	655	13	BH097469	BH097469 RPCI-24-3
c	5	16.4	91.1	155	10	AW346676	AW346676 29231 MAR
c	6	16.4	91.1	166	10	AW346655	AW346655 29171 MAR
c	7	16.4	91.1	232	10	AV312947	AV312947 AV312947
c	8	16.4	91.1	254	10	AV289929	AV289929 AV289929
	9	16.4	91.1	261	11	BI184980	BI184980 UNL-P-FN-
	10	16.4	91.1	290	13	AZ113096	AZ113096 RPCI-23-4
	11	16.4	91.1	294	11	BF713588	BF713588 MI-P-O2-a
	12	16.4	91.1	300	11	BI184279	BI184279 UNL-P-FN-
c	13	16.4	91.1	370	13	AZ313323	AZ313323 1M0029A06
	14	16.4	91.1	379	13	AZ793176	AZ793176 2M0046N11
	15	16.4	91.1	395	13	AZ831760	AZ831760 2M0111K09
c	16	16.4	91.1	425	13	AZ036915	AZ036915 RPCI-23-3
	17	16.4	91.1	427	13	AZ558978	AZ558978 RPCI-23-1
	18	16.4	91.1	432	13	AZ909325	AZ909325 RPCI-24-1
	19	16.4	91.1	458	11	BI185480	BI185480 UNL-P-FN-
	20	16.4	91.1	468	11	BI182727	BI182727 UNL-P-FN-
	21	16.4	91.1	468	11	BI186066	BI186066 UNL-P-FN-
c	22	16.4	91.1	470	13	AZ387701	AZ387701 1M0147M17
	23	16.4	91.1	495	13	AZ955800	AZ955800 2M0222K05
	24	16.4	91.1	499	11	BF704354	BF704354 MI-P-O3-a
	25	16.4	91.1	513	10	AW786160	AW786160 119054 MA
	26	16.4	91.1	521	11	BI399852	BI399852 MI-P-AY1-
	27	16.4	91.1	552	13	AZ242483	AZ242483 RPCI-23-7
	28	16.4	91.1	553	13	AZ999118	AZ999118 2M0286B21
	29	16.4	91.1	579	11	BI183723	BI183723 UNL-P-FN-
	30	16.4	91.1	602	11	BF712162	BF712162 MI-P-E6-a
	31	16.4	91.1	607	13	BH088353	BH088353 RPCI-24-2
c	32	16.4	91.1	613	13	AZ454083	AZ454083 1M0255F20
c	33	16.4	91.1	615	13	AZ385535	AZ385535 1M0144M06
	34	16.4	91.1	622	13	AZ115761	AZ115761 RPCI-23-4
	35	16.4	91.1	655	13	AZ836447	AZ836447 2M0131J13
	36	16.4	91.1	681	13	AZ115038	AZ115038 RPCI-23-4
	37	16.4	91.1	683	11	BG122508	BG122508 602353118
c	38	16.4	91.1	767	13	BH125701	BH125701 RPCI-24-3
c	39	16.4	91.1	864	11	BF790922	BF790922 602250949
c	40	16.4	91.1	1042	10	BE544440	BE544440 601077062
	41	16.4	91.1	2260	12	AK015210	AK015210 Mus muscu
	42	16	88.9	686	11	BF501296	BF501296 AT16580.5
	43	15.4	85.6	142	13	AZ601453	AZ601453 1M0419B14
c	44	15.4	85.6	252	11	W66405	W66405 TgESTzy93e0
c	45	15.4	85.6	259	13	AZ306532	AZ306532 1M0007I13

ALIGNMENTS

RESULT 1
BI360879/c
LOCUS BI360879 558 bp mRNA EST 01-AUG-2001
DEFINITION 388957 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BI360879
VERSION BI360879.1 GI:15056907
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 558)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W. and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
PCR PRimers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 142 row: B column: 4
Seq primer: ATTTAGGTGACACTATAG.
FEATURES Location/Qualifiers
source 1. .558
/organism="Sus scrofa"

```
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      129 a    123 c    119 g    187 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 11;  Length 558;
  Best Local Similarity  100.0%;  Pred. No. 76;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          ||||||||||||||||||
Db    178 CCACAGTCGTACACCACG 161


RESULT   2
AZ056887
LOCUS       AZ056887      323 bp    DNA             GSS       30-MAR-2000
DEFINITION  RPCI-23-428K7.TJ RPCI-23 Mus musculus genomic clone RPCI-23-428K7,
            DNA sequence.
ACCESSION   AZ056887
VERSION     AZ056887.1  GI:7348123
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 323)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 428  row: K  column: 7
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .323
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-428K7"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       83 a     59 c     72 g    109 t
ORIGIN


  Query Match           94.4%;  Score 17;  DB 13;  Length 323;
  Best Local Similarity  100.0%;  Pred. No. 2.1e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccac 17
          |||||||||||||||||
Db    218 CCACAGTCGTACACCAC 234


RESULT   3
AZ749422/c
LOCUS       AZ749422      559 bp    DNA             GSS       25-JAN-2001
DEFINITION  RPCI-24-115E19.TV RPCI-24 Mus musculus genomic clone RPCI-24-115E19
            , DNA sequence.
ACCESSION   AZ749422
VERSION     AZ749422.1  GI:12534581
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-115E19.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 115  row: E  column: 19
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-115E19"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      133 a    125 c    137 g    164 t
ORIGIN


  Query Match           94.4%;  Score 17;  DB 13;  Length 559;
  Best Local Similarity  100.0%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccac 17
          |||||||||||||||||
Db    411 CCACAGTCGTACACCAC 395


RESULT   4
BH097469
```

```
LOCUS       BH097469      655 bp    DNA             GSS       19-JUL-2001
DEFINITION  RPCI-24-315L22.TV RPCI-24 Mus musculus genomic clone RPCI-24-315L22
            , DNA sequence.
ACCESSION   BH097469
VERSION     BH097469.1  GI:14918815
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 655)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-315L22.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 315  row: L  column: 22
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .655
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-315L22"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      180 a    157 c    148 g    170 t
ORIGIN


  Query Match             94.4%;  Score 17;  DB 13;  Length 655;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccac 17
          |||||||||||||||||
Db    116 CCACAGTCGTACACCAC 132


RESULT   5
AW346676/c
LOCUS       AW346676      155 bp    mRNA            EST       09-JUL-2000
DEFINITION  29231 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   AW346676
VERSION     AW346676.1  GI:6844386
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 155)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 13  row: B  column: 12
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .155
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT       34 a     45 c     39 g     37 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 10;  Length 155;
  Best Local Similarity   94.4%;  Pred. No. 3.5e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          |||||||||||| |||||
Db    112 CCACAGTCGTACGCCACG 95


RESULT   6
AW346655/c
LOCUS       AW346655      166 bp    mRNA            EST       09-JUL-2000
DEFINITION  29171 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   AW346655
VERSION     AW346655.1  GI:6844365
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 166)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 13  row: H  column: 6
            Seq primer: ATTTAGGTGACACTATAG.
```

```
FEATURES             Location/Qualifiers
     source          1. .166
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT       45 a     45 c     39 g     37 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 10;  Length 166;
  Best Local Similarity   94.4%;  Pred. No. 3.5e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          |||||||||||| |||||
Db    112 CCACAGTCGTACGCCACG 95


RESULT   7
AV312947/c
LOCUS       AV312947      232 bp    mRNA            EST       08-NOV-1999
DEFINITION  AV312947 RIKEN full-length enriched, adult male thymus Mus musculus
            cDNA clone 5830407B15 3', mRNA sequence.
ACCESSION   AV312947
VERSION     AV312947.1  GI:6278199
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 232)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .232
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5830407B15"
                     /clone_lib="RIKEN full-length enriched, adult male thymus"
                     /sex="male"
                     /tissue_type="thymus"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI."
BASE COUNT       49 a     64 c     39 g     80 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 10;  Length 232;
  Best Local Similarity   94.4%;  Pred. No. 3.8e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          ||||| ||||||||||||
Db    190 CCACATTCGTACACCACG 173


RESULT   8
AV289929/c
LOCUS       AV289929      254 bp    mRNA            EST       09-NOV-1999
DEFINITION  AV289929 RIKEN full-length enriched, 11 days pregnant adult female
            ovary and uterus Mus musculus cDNA clone 5033430G19, mRNA sequence.
ACCESSION   AV289929
VERSION     AV289929.1  GI:6303960
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 254)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
```

```
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .254
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5033430G19"
                     /clone_lib="RIKEN full-length enriched, 11 days pregnant
                     adult female ovary and uterus"
                     /sex="female"
                     /tissue_type="ovary and uterus"
                     /dev_stage="11 days pregnant, adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI."
BASE COUNT       51 a     74 c     46 g     83 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 10;  Length 254;
  Best Local Similarity   94.4%;  Pred. No. 3.9e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          |||||||| |||||||||
Db    212 CCACAGTCATACACCACG 195


RESULT   9
BI184980
LOCUS       BI184980      261 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-dc-h-12-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-dc-h-12-0-UNL 3', mRNA sequence.
ACCESSION   BI184980
VERSION     BI184980.1  GI:14659389
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized porcine ovarian follicles library The following
            repetitive elements were found in this cDNA sequence: 255-304,
            >AT_rich#Low_complexity
            Seq primer: M13 -29
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .261
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-dc-h-12-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_LIB=UNL-P-FN
                     TAG_TISSUE=porcine ovarian follicles
                     TAG_SEQ=CACACT"
BASE COUNT       64 a     52 c     66 g     78 t      1 others
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 11;  Length 261;
  Best Local Similarity   94.4%;  Pred. No. 3.9e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          |||||||| |||||||||
Db     64 CCACAGTCATACACCACG 81


RESULT  10
AZ113096
LOCUS       AZ113096      290 bp    DNA             GSS       12-MAY-2000
DEFINITION  RPCI-23-475F22.TJ RPCI-23 Mus musculus genomic clone RPCI-23-475F22
            , DNA sequence.
ACCESSION   AZ113096
VERSION     AZ113096.1  GI:7772367
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 290)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
```

```
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-475F22.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 475  row: F  column: 22
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .290
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-475F22"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       84 a     59 c     84 g     63 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 13;  Length 290;
  Best Local Similarity   94.4%;  Pred. No. 4e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccacg 18
           |||||||| |||||||||
Db     191 CCACAGTCATACACCACG 208


RESULT  11
BF713588
LOCUS       BF713588      294 bp    mRNA            EST       02-JAN-2001
DEFINITION  MI-P-O2-adx-e-01-1-UM.s1 MI-P-O2 Sus scrofa cDNA clone
            MI-P-O2-adx-e-01-1-UM 3', mRNA sequence.
ACCESSION   BF713588
VERSION     BF713588.1  GI:12013063
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: RJ Woods, JA Green, RS Prather S142
            Animal Science Research Center, Department of Animal Science,
            University of Missouri-Columbia, 65211 Clone distribution: clones
            will be available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .294
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-O2-adx-e-01-1-UM"
                     /clone_lib="MI-P-O2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-O2
                     library is derived from ovary at estrus day 5.  For a
                     detailed description of the  library from which this clone
                     was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/.
                     TAG_SEQ=None found"
BASE COUNT       69 a     61 c     77 g     87 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 11;  Length 294;
  Best Local Similarity   94.4%;  Pred. No. 4e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ccacagtcgtacaccacg 18
           |||||||| |||||||||
Db      62 CCACAGTCATACACCACG 79


RESULT  12
BI184279
LOCUS       BI184279      300 bp    mRNA            EST       10-JUL-2001
DEFINITION  UNL-P-FN-ck-g-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
            UNL-P-FN-ck-g-02-0-UNL 3', mRNA sequence.
ACCESSION   BI184279
VERSION     BI184279.1  GI:14658688
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Caetano,A.R., Johnson,R.K. and Pomp,D.
  TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail.
            Seq primer: M13 -29
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Sus scrofa"
                     /strain="University of Nebraska, Lincoln Swine Selection
                     Lines"
```

```
                     /db_xref="taxon:9823"
                     /clone="UNL-P-FN-ck-g-02-0-UNL"
                     /clone_lib="UNL-P-FN"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
                     library is a normalized library representing porcine
                     ovarian follicles, ranging between 2.0 to 10.0 mm in
                     diameter, collected  during 7 days of the follicular phase
                     of the pig estrous cycle. This  library was derived from
                     the library UNL-P-F2. The tag is a string of  5-6
                     nucleotides present between the Not I site and the
                     oligo-dT track.  The library was constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996.
                     TAG_SEQ=None found"
BASE COUNT       69 a     62 c     71 g     97 t      1 others
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 11;  Length 300;
  Best Local Similarity   94.4%;  Pred. No. 4e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          |||||||| |||||||||
Db     59 CCACAGTCATACACCACG 76


RESULT  13
AZ313323/c
LOCUS       AZ313323      370 bp    DNA              GSS       29-SEP-2000
DEFINITION  1M0029A06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0029A06 R, DNA sequence.
ACCESSION   AZ313323
VERSION     AZ313323.1  GI:10358137
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0029  row: A  column: 06
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 370.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0029A06"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       66 a    100 c     84 g    120 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 13;  Length 370;
  Best Local Similarity   94.4%;  Pred. No. 4.2e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccacagtcgtacaccacg 18
          |||| ||||||||||||||
Db    165 CCACGGTCGTACACCACG 148


RESULT  14
AZ793176
LOCUS       AZ793176      379 bp    DNA              GSS       16-FEB-2001
DEFINITION  2M0046N11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0046N11 F, DNA sequence.
ACCESSION   AZ793176
VERSION     AZ793176.1  GI:12937866
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0046  row: N  column: 11
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 379.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0046N11"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
```

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 120 a 80 c 95 g 84 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 379;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccacagtcgtacaccacg 18
||||||||| |||||||||
Db 75 CCACAGTCATACACCACG 92

RESULT 15
AZ831760
LOCUS AZ831760 395 bp DNA GSS 20-FEB-2001
DEFINITION 2M0111K09R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0111K09 R, DNA sequence.
ACCESSION AZ831760
VERSION AZ831760.1 GI:13001668
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 395)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0111 row: K column: 09
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 395.
FEATURES Location/Qualifiers
source 1. .395
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0111K09"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 121 a 88 c 112 g 74 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 395;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccacagtcgtacaccacg 18
||||||||| |||||||||
Db 253 CCACAGTCATACACCACG 270

Search completed: February 23, 2002, 23:37:23
Job time: 9444 sec